

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2004, 18:27:09 ; Search time 46 seconds

(without alignments)  
4492.705 Million cell updates/sec

Title: US-09-961-086-1

Perfect score: 3352

Sequence: 1 MSSNNVEVPIPVQSGNTNGF.....MIVIFLTAYLKILFLKKYS 655

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_virus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3346	99.8	655	4 Q96TA8	Q96TA8 homo sapien
2	3342	99.7	655	4 Q81X16	Q81X16 homo sapien
3	3338	99.6	655	4 Q96LD6	Q96LD6 homo sapien
4	2849.5	95.0	656	6 Q8MIB3	Q8MIB3 sus scrofa
5	2762	82.4	657	11 Q7TMS5	Q7TMS5 mus musculus
6	2761	82.4	541	4 Q86V64	Q86V64 homo sapien
7	2757	82.2	657	11 Q9R004	Q9R004 mus musculus
8	2754	82.2	657	11 Q8W57	Q8W57 rattus norv
9	2752	82.1	657	11 Q80ST1	Q80ST1 rattus norv
10	2742	81.8	657	11 Q80XF3	Q80XF3 rattus norv
11	1787.5	53.3	650	11 Q8KIS	Q8KIS mus musculus
12	847.5	25.3	801	5 Q8T691	Q8T691 dictyosteli
13	809	24.1	751	10 Q93YS4	Q93YS4 arabidopsis
14	808	24.1	687	5 Q9NH94	Q9NH94 bombyx mori
15	806	24.0	687	5 Q94960	Q94960 drosophila
16	800.5	23.9	737	10 Q9FT51	Q9FT51 arabidopsis

17	793	23.7	679	5 Q9BH97	Q9BH97 ceratitis c
18	790	23.6	692	5 P91892	P91892 aedes aegypt
19	783.5	23.4	648	10 Q9C6W5	Q9C6W5 arabidopsis
20	777.5	23.2	679	5 Q8IS30	Q8IS30 bacterocera
21	776.5	23.2	567	10 Q9FG17	Q9FG17 arabidopsis
22	774	23.1	646	10 Q9C6R7	Q9C6R7 arabidopsis
23	772	23.1	670	5 Q77423	Q77423 bacterocera
24	767.5	22.9	687	10 Q9C8K2	Q9C8K2 arabidopsis
25	766.5	22.9	695	10 Q8LMQ5	Q8LMQ5 oryza sativ
26	763.5	22.8	798	5 Q8T689	Q8T689 dictyosteli
27	759	22.6	703	10 Q8RXN0	Q8RXN0 arabidopsis
28	755	22.5	672	10 Q9LI82	Q9LI82 arabidopsis
29	751	22.4	669	5 Q8WRE2	Q8WRE2 tribolium c
30	749.5	22.4	691	10 Q8RWI9	Q8RWI9 arabidopsis
31	746	22.3	649	10 Q9SIT6	Q9SIT6 arabidopsis
32	745	22.2	725	10 Q9M3D6	Q9M3D6 arabidopsis
33	744	22.2	669	5 Q8WRI1	Q8WRI1 tribolium c
34	743.5	22.2	1528	5 Q8T677	Q8T677 dictyosteli
35	743	22.2	1328	5 Q9NGP5	Q9NGP5 dictyosteli
36	741.5	22.1	739	10 Q9LFG8	Q9LFG8 arabidopsis
37	739	22.0	678	10 Q9C8J8	Q9C8J8 arabidopsis
38	739	22.0	678	10 Q8GX48	Q8GX48 arabidopsis
39	730.5	21.8	662	10 Q949Y4	Q949Y4 arabidopsis
40	730.5	21.8	662	10 Q84TH5	Q84TH5 arabidopsis
41	730.5	21.8	1520	5 Q8T687	Q8T687 dictyosteli
42	730	21.8	626	5 Q8T684	Q8T684 dictyosteli
43	730	21.8	626	5 Q86HQ2	Q86HQ2 dictyosteli
44	728.5	21.7	692	10 Q7XUM2	Q7XUM2 oryza sativ
45	727.5	21.7	668	10 Q9ARU4	Q9ARU4 oryza sativ

## ALIGNMENTS

### RESULT 1

Q96TA8 ID Q96TA8 PRELIMINARY; PRT; 655 AA.

AC Q96TA8;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE ATP-binding cassette superfamily G (White) member 2 (Hypothetical protein).  
DE ABCG2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=2120193; PubMed=11306452;

RA Komatani H., Kotani H., Hara Y., Nakagawa R., Matsumoto M.,

RA Arakawa H., Nishimura S.;

RT "Identification of breast cancer resistant protein/mitoxantrone

RT resistance/placenta-specific, ATP-binding cassette transporter as a

RT transporter of NB-506 and J-107088, topoisomerase I inhibitors with an

RT indolocarbazole structure.";

RL Cancer Res. 61:2827-2832(2001).

RN [2]

RP SEQUENCE FROM N.A.

RX TISSUE=Fancreatic carcinoma;

RA Strausberg R.;

RA Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.

EMBL; ABC051855; BAB46933.1; -

EMBL; BC021281; AAH21281.1; -

DR GO; GO:0016020; C:membrane; IEA.

DR GO; GO:0005524; F:ATP binding; IEA.

DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.

DR GO; GO:0008810; P:transporter; IEA.

DR InterPro; IPR003439; ABC\_transporter.

DR Pfam; PF00005; ABC\_tran; 1.

DR ProDom; PD000006; ABC\_transporter; 1.



OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Schuetz J.D., Wall A.M., Sampath J., Sorrentino B., Du G.;  
 RT "The Human ABC Transporter, ABCG2, Transports Hoechst 33342 and  
 RL Requires an Intact Walker A Motif."  
 DR ENBL, AY017168; AAG52982.1; -  
 DR GO:0016020; C:membrane; IEA.  
 DR GO:0005524; F:ATP binding; IEA.  
 DR GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.  
 DR GO:0000166; F:nucleotide binding; IEA.  
 DR GO:0006810; P:transport; IEA.  
 DR InterPro: IPR003593; AAA\_ATPase.  
 DR InterPro: IPR003439; ABC\_transporter.  
 DR InterPro: IPR006162; Pfamne\_S.  
 DR Pfam: PF00005; ABC\_tran; 1.  
 DR ProDom: PD000006; ABC\_transporter; 1.  
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.  
 KW ATP-binding.  
 SQ SEQUENCE 555 AA; 72288 MW; B3B5DC02C095C4A8 CRC64;  
 Query Match 99.6%; Score 3338; DB 4; Length 555;  
 Best Local Similarity 99.7%; Pred. No. 5.1e-225;  
 Matches 653; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MSSSNVEFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICVYVKLSGFLPCRKVE 60  
 DB 1 MSSSNVEFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICVYVKLSGFLPCRKVE 60  
 QY 61 KEILSNITMGKPLNAILGPTGGKSSLLDLAARKDPSGLSGDVLINGAPRANPKCN 120  
 DB 61 KEILSNITMGKPLNAILGPTGGKSSLLDLAARKDPSGLSGDVLINGAPRANPKCN 120  
 QY 121 SGVYQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIQELGLDKVADSKVGT 180  
 DB 121 SGVYQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIQELGLDKVADSKVGT 180  
 QY 181 QFIRGVSGGERKRTSGMELITDPSILFDEPTTGLDSTANAVLLLLKMSKQGRITIF 240  
 DB 181 QFIRGVSGGERKRTSGMELITDPSILFDEPTTGLDSTANAVLLLLKMSKQGRITIF 240  
 QY 241 SIHOPRYSIFKLFDLSLTLASGLMFGHPAQALGYFESAGVHCEAYNNPADFFLDIING 300  
 DB 241 SIHOPRYSIFKLFDLSLTLASGLMFGHPAQALGYFESAGVHCEAYNNPADFFLDIING 300  
 QY 301 DSTAVANREDEKATEIIPSKODKPLIEKLAETVNSFFYKETAELHOLSGGEKKK 360  
 DB 301 DSTAVANREDEKATEIIPSKODKPLIEKLAETVNSFFYKETAELHOLSGGEKKK 360  
 QY 361 ITVPKEISYTTSFCHOLRWYSKRSFKNLGNPOASIAQIIIVTVGLVIGAIYFGLKND 420  
 DB 361 ITVPKEISYTTSFCHOLRWYSKRSFKNLGNPOASIAQIIIVTVGLVIGAIYFGLKND 420  
 QY 421 TGIONRAGVLPFLTTCQPSVSAVELFVVEKKLFIEHISGYRVSFVFLGLSLDLLP 480  
 DB 421 TGIONRAGVLPFLTTCQPSVSAVELFVVEKKLFIEHISGYRVSFVFLGLSLDLLP 480  
 QY 481 MTMLPSIIFTCIVVMGLKPKADAFVMMFTLMWVAYSASSMALAIAAGQSVSVATLL 540  
 DB 481 MRMLPSIIFTCIVVMGLKPKADAFVMMFTLMWVAYSASSMALAIAAGQSVSVATLL 540  
 QY 541 MTICFVPMWIFSGLLVNLITIASLWQLVFSIPRYGFTALQHNFCPLGNATGN 600  
 DB 541 MTICFVPMWIFSGLLVNLITIASLWQLVFSIPRYGFTALQHNFCPLGNATGN 600  
 QY 601 NPCNVATCTGEBVLVKQIDLSFWGLWKNHVALACMIVFLTIAYLKLFLKKYS 655

DB 601 NPCNVATCTGEBVLVKQIDLSFWGLWKNHVALACMIVFLTIAYLKLFLKKYS 655  
 RESULT 4  
 QSMIB3 PRELIMINARY; PRT; 656 AA.  
 ID QSMIB3;  
 AC QSMIB3;  
 DT 01-OCT-2002 (Tremblrel. 22, Created)  
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE Brain multidrug resistance protein.  
 GN BMDP.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22050127; PubMed=12054514;  
 RA Eisenblatter T., Galla H.J.;  
 RT "A new multidrug resistance protein at the blood-brain barrier."  
 RL Biochem. Biophys. Res. Commun. 293:1273-1278(2002).  
 DR ENBL; AJ420927; CAD12785.1; -  
 DR PIR; JC7860; JC7860.  
 DR GO:0016020; C:membrane; IEA.  
 DR GO:0005524; F:ATP binding; IEA.  
 DR GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.  
 DR GO:0000166; F:nucleotide binding; IEA.  
 DR GO:0006810; P:transport; IEA.  
 DR InterPro: IPR003593; AAA\_ATPase.  
 DR InterPro: IPR003439; ABC\_transporter.  
 DR Pfam: PF00005; ABC\_tran; 1.  
 DR ProDom: PD000006; ABC\_transporter; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR PROSITE; PS00893; ABC\_TRANSPORTER\_2; 1.  
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.  
 KW ATP-binding.  
 SQ SEQUENCE 656 AA; 72392 MW; 118ADD5B53D9D67F CRC64;  
 Query Match 85.0%; Score 2849.5; DB 6; Length 656;  
 Best Local Similarity 84.3%; Pred. No. 4e-189;  
 Matches 553; Conservative 44; Mismatches 58; Indels 1; Gaps 1;  
 QY 1 MSSSNVEFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICVYVKLSGFLPCRKVE 60  
 DB 1 MSSSNVEFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICVYVKLSGFLPCRKVE 60  
 QY 61 KEILSNITMGKPLNAILGPTGGKSSLLDLAARKDPSGLSGDVLINGAPRANPKCN 120  
 DB 61 KEILSNITMGKPLNAILGPTGGKSSLLDLAARKDPSGLSGDVLINGAPRANPKCN 120  
 QY 121 SGVYQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIQELGLDKVADSKVGT 180  
 DB 121 SGVYQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIQELGLDKVADSKVGT 180  
 QY 181 QFIRGVSGGERKRTSGMELITDPSILFDEPTTGLDSTANAVLLLLKMSKQGRITIF 240  
 DB 181 QFIRGVSGGERKRTSGMELITDPSILFDEPTTGLDSTANAVLLLLKMSKQGRITIF 240  
 QY 241 SIHOPRYSIFKLFDLSLTLASGLMFGHPAQALGYFESAGVHCEAYNNPADFFLDIING 300  
 DB 241 SIHOPRYSIFKLFDLSLTLASGLMFGHPAQALGYFESAGVHCEAYNNPADFFLDIING 300  
 QY 301 DSTAVANREDEKATEIIPSKODKPLIEKLAETVNSFFYKETAELHOLSGGEKKK 359  
 DB 301 DSAAVLSRADDEGAQEPPEPEKOTFLIDKLAAFYTNSSFFKDTKVELDQSGGRRKK 360  
 QY 360 KITVPKEISYTTSFCHOLRWYSKRSFKNLGNPOASIAQIIIVTVGLVIGAIYFGLKND 419  
 DB 361 KSSVYKEVYTTTSFCHOLRWISRSFKNLGNPOASIAQIIIVTVGLVIGAIYFGLKND 420  
 QY 420 STGIQNRAGVLPFLTTCQPSVSAVELFVVEKKLFIEHISGYRVSFVFLGLSLDLL 479

```

421 PSQIQNRAGVLFELTNNQCFSSVSAVELLVVEKLFIEHYISGYRVSFFGKLLSDLL 480
480 PWTMLPSIIFTCITVYFMGLGKPKADAFVMMFTLMMVAYSASSMALAIAAGOSVVSATL 539
481 PWRMLPSIIFTCITVYFMGLGKPKAVGSFFMMFTLMMVAYSASSMALAIAAGOSVVSATL 540
540 LMTICFVFMFISGLLVNLTITIASMLSWLOYSIPRYGFTALOHNEFLGQFCPLGNATG 599
541 LMTISFVFMFISGLLVNLTITVPMLSWLOYSIPRYGFTALOHNEFLGQFCPLGNATG 600
600 NNPNCVATCTGEEYLVKQGDIDSPMGLWKNHVALACMIVIFLTIAVLKLLFLKKYS 655
601 NNTCSFAICTGAEYLENGISLSAWGLWQNHVALACMIVIFLTIAVLKLLFLKKYS 656

RESULT 5
Q7TMS5
ID Q7TMS5 PRELIMINARY; PRT; 657 AA.
AC Q7TMS5;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE ATP-binding cassette, sub-family G, member 2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6NCR; TISSUE=Hematopoietic Stem Cell;
RX MEDLINE=22398257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Richards S., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6NCR; TISSUE=Hematopoietic Stem Cell;
RA Strausberg R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC053730; ARL53730.1; -.
RW ATP-binding.
SQ SEQUENCE 657 AA; 72977 MW; DCD70CSD9FA2BA5F CRC64;

Query Match 82.4%; Score 2762; DB 11; Length 657;
Best Local Similarity 81.5%; Pred. No. 4.7e-183;
Matches 536; Conservative 52; Mismatches 56; Indels 4; Gaps 3;

QY 1 MSSNNVEFIPVSGNGTNGPPATASNDLKAFTEGAVLSFNICVYKVLKSGFLPCRKYE 60
Db 1 MSSNDHVLVPMSSQNNGLPRTNSRAVRTLAEGDVLSPFHITVYKVKSGFL-VRKTYE 59
QY 61 KEILSNINGIMKGLNAILGPTGGKSSLLDVLAAKDPGSLGSDVLINGAPPAFKN 120
Db 60 KEILSDINGIMKGLNAILGPTGGKSSLLDVLAAKDPKGLSGDVLINGAPPAFKN 119

```

```

QY 121 SGVVQDDVVMGTLTIVRENLQPSAALRLATTTNHEKNERINRVIOELGDKVADSKVGT 180
Db 120 SGVVQDDVVMGTLTIVRENLQPSAALRLPTTKNHEKNERINTIKELGLEKVDKSKVGT 179
QY 181 QFIRGVSGGERKRTSICMELITDPSILFLDEPTTGLDSDSTANAVLLLLKRMKSKQRTIIF 240
Db 180 QFIRGISGGERKRTSICMELITDPSILFLDEPTTGLDSDSTANAVLLLLKRMKSKQRTIIF 239
QY 241 SIHQPRYSIPKLPDSITLTLASGLMHFGPAQOALGYFESAGYHCEAYNNPADPFLDLING 300
Db 240 SIHQPRYSIPKLPDSITLTLASGLVPHGPAQKALEYFASAGYHCEBYNNPADPFLDLING 299
QY 301 DSTAVALNREE-DFKATEIETPSKQDKPLIEKLAIIYVNSFFYKTKAELHQLSGGKKK 359
Db 300 DSAVMLNREEQDNEANKTEPSKQKPVLENSEFYNSAIYGETKAELEDQPGAQEKK 359
QY 360 KITVFKESYTTSCFQHLRWKSRSPKLLGNPQASIAQIIVTVVLGLVIGALYFGLKND 419
Db 360 GTSAFKEPVYVTSFCHQLRWIARRSPKLLGNPQASVAQIIVTVLGLIIGALYFGLKYD 419
QY 420 STGIQNEAGVLFELTNNQCFSSVSAVELFVVEKLFIEHYISGYRVSFFGKLLSDLL 479
Db 420 AAGQNEAGVLFELTNNQCFSSVSAVELFVVEKLFIEHYISGYRVSFFGKLLSDLL 479
QY 480 PWTMLPSIIFTCITVYFMGLGKPKADAFVMMFTLMMVAYSASSMALAIAAGOSVVSATL 539
Db 480 PWRMLPSIIFTCITVYFMGLGKPKAVGSFFMMFTLMMVAYSASSMALAIAAGOSVVSATL 539
QY 540 LMTICFVFMFISGLLVNLTITIASMLSWLOYSIPRYGFTALOHNEFLGQFCPLGNATG 599
Db 540 LMTISFVFMFISGLLVNLTITVPMLSWLOYSIPRYGFTALOHNEFLGQFCPLGNATG 599
QY 600 NNPNC-VYATCTGEEYLVKQGDIDSPMGLWKNHVALACMIVIFLTIAVLKLLFLKKYS 655
Db 600 NSTCVNSVAICTGNEYLINQIELSIPMGLWKNHVALACMIIIFLTIAVLKLLFLKKYS 657

RESULT 6
Q86V64
ID Q86V64 PRELIMINARY; PRT; 541 AA.
AC Q86V64;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE ATP-binding cassette sub-family G member 2 (Fragment).
GN ABCG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RP Zhang W., Petrovic J.-M., Andrade M.F., Zhang H., Stanimirovic D.B.;
RA "The Expression and Functional Characterization of ABCG2 in Brain
RT Endothelial Cells and Vessels."
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY288307; AAP31310.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR006162; Ppantre_S.
DR Pfam; PF000005; ABC_tran; 1.
DR ProDom; PDC00006; ABC_transporter; 1.
DR PROSITE; PS00012; PROSHOPANTHETHEINE; 1.
KW ATP-binding.
FT NON_TER
SQ SEQUENCE 541 AA; 60450 MW; 050B2742C6A3F66C CRC64;

Query Match 82.4%; Score 2761; DB 4; Length 541;
Best Local Similarity 99.8%; Pred. No. 4.2e-183;
Matches 540; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```



```

Qy 115 ANFKCSGYVQDDVVMGTLTVRENLQFSAALRLATTMTNHEKNERINRVIOELGLDKA 174
Db 1 ANFKCSGYVQDDVVMGTLTVRENLQFSAALRLATTMTNHEKNERINRVIOELGLDKA 60
Qy 175 DSKVGTGIFRGVSGGKRTSICWELITDPSILFDEPTTGDDSTANAVLLLLKRMGSKQ 234
Db 61 DSKVGTGIFRGVSGGKRTSICWELITDPSILFDEPTTGDDSTANAVLLLLKRMGSKQ 120
Qy 235 GRTIISIHOPRYSIKLFDLSLTLASGRMLFHPGPAQALGYFESAGYHCEAYNPPDF 294
Db 121 GRTIISIHOPRYSIKLFDLSLTLASGRMLFHPGPAQALGYFESAGYHCEAYNPPDF 180
Qy 295 LDIINGDSTAVLNREEDFKATEIIEPSKQDKPLIEKLAIEIYVNSPFYKETAELHQLSG 354
Db 181 LDIINGDSTAVLNREEDFKATEIIEPSKQDKPLIEKLAIEIYVNSPFYKETAELHQLSG 240
Qy 355 GEKKKKTIVPEISYTSFCHQRLRWYSKSFKNLGNPOASTAOIIVTVVLGIVGAIYF 414
Db 241 GEKKKKTIVPEISYTSFCHQRLRWYSKSFKNLGNPOASTAOIIVTVVLGIVGAIYF 300
Qy 415 GLKNDSTGIQNRAGVLFLLTNNOCFSSVSAVELFVVEKKLFTHIYISGYRVSSYFLGKL 474
Db 301 GLKNDSTGIQNRAGVLFLLTNNOCFSSVSAVELFVVEKKLFTHIYISGYRVSSYFLGKL 360
Qy 475 LSDLLPMTLPSIIFTCIVVFMGLKPKADAFVVMFTLMVAYSSSMALAIAGQSVV 534
Db 361 LSDLLPMTLPSIIFTCIVVFMGLKPKADAFVVMFTLMVAYSSSMALAIAGQSVV 420
Qy 535 SVATLLMTICFVFMFIFSGLLVNLTTIASLWQYFSIPRYGFTALQHNFPNGFCG 594
Db 421 SVATLLMTICFVFMFIFSGLLVNLTTIASLWQYFSIPRYGFTALQHNFPNGFCG 480
Qy 595 LNATGNPNVACTGEEYLVKQIDLSPWGLWKNHVALACMIVIFLTAYLKLFLKKY 654
Db 481 LNATGNPNVACTGEEYLVKQIDLSPWGLWKNHVALACMIVIFLTAYLKLFLKKY 540
Qy 655 S 655
Db 541 S 541

RESULT 7
Q9R004 PRELIMINARY; PRT; 657 AA.
ID Q9R004
AC Q9R004;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Breast cancer resistance protein 1.
GN ABCG2 OR BCRP1
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10080;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB; TISSUE=Liver;
RX MEDLINE=99413474; PubMed=10485464;
RA Allen J.D., Brinkhuis R.F., Wijnholds J., Schinkel A.H.;
RT "The mouse Bcrp/Mxr/Abcg Gene: amplification and overexpression in
RT cell lines selected for resistance to topotecan, mitoxantrone, or
RT doxorubicin."
RL Cancer Res. 59:4237-4241(1999).
DR EMBL; AF140218; AAD54216.1; -.
DR GDB; MGI:1347061; Abcg2.
DR GO; GO:0016020; C.membrane; IEA.
DR GO; GO:0005524; F.ATP binding; IEA.
DR GO; GO:0004009; F.ATP-binding cassette (ABC) transporter acti. .; IEA.
DR GO; GO:0000166; F.nucleotide binding; IEA.
DR GO; GO:0006810; P.transport; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.

```

```

InterPro; IPR006162; Ppantne_S.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00893; ABC TRANSPORTER 2; 1.
DR PROSITE; PS00012; PHOSPHOPANTHETINE; 1.
KW ATP-binding.
SQ SEQUENCE 657 AA; 73021 MW; 207B70BC272CC0D5 CRC64;

Query Match 82.2%; Score 2757; DB 11; Length 657;
Best Local Similarity 81.5%; Pred. No. 1e-182;
Matches 536; Conservative 51; Mismatches 67; Indels 4; Gaps 3;

Qy 1 MSSNNVEFIPYVQSGNTNGFPATASNDLKAFTEGAVLSFHNICRYRVLKSGFLPCRPVE 60
Db 1 MSSNDHVLVPMSQRNNNGLPMSRAVRLTLAGDVLSPFHTYRVKYSGL-VRKIVE 59
Qy 61 KEILSNINIMKPGNALILPTGGGKSLLDVLAARKDPSGLSDVLINGAPRANPKCN 120
Db 60 KEILSDINIMKPGNALILPTGGGKSLLDVLAARKDPSGLSDVLINGAPQPAHFKCC 119
Qy 121 SGYVQDDVVMGTLTVRENLQFSAALRLATTMTNHEKNERINRVIOELGLDKVADSKVGT 180
Db 120 SGYVQDDVVMGTLTVRENLQFSAALRLATTMTNHEKNERINTIIEKLEKADSKVGT 179
Qy 181 QTRGVSGGERKRTSIGMELITDPSILFDEPTTGDDSTANAVLLLLKRMKSGKRTIIF 240
Db 180 QTRGVSGGERKRTSIGMELITDPSILFDEPTTGDDSTANAVLLLLKRMKSGKRTIIF 239
Qy 241 SIHQPRYSIFKLFDLSLTLASGRMLFHPGPAQALGYFESAGYHCEAYNPPADFFLDING 300
Db 240 SIHQPRYSIFKLFDLSLTLASGRMLFHPGPAQALGYFESAGYHCEAYNPPADFFLDING 299
Qy 301 DSTAVALARBE-DFKATEIIEPSKQDKPLIEKLAIEIYVNSPFYKETAELHQLSGGKK 359
Db 300 DSAVNLNEEDQNEANKTEPESKGEKPEIENLSEFYINSIYGETKAELDQLPQAEKK 359
Qy 360 KITVFKESYTSFCHQRLRWYSKSFKNLGNPOASTAOIIVTVVLGIVGAIYFGLKND 419
Db 360 GT-SAPKEPVYTSFCHQRLRWYARRSFKNLGNPOASTAOIIVTVVLGIVGAIYFGLKYD 419
Qy 420 STGIONRAGVLFLLTNNOCFSSVSAVELFVVEKKLFTHIYISGYRVSSYFLGKLSDLL 479
Db 420 AAGMQRAGVLFLLTNNOCFSSVSAVELFVVEKKLFTHIYISGYRVSSYFLGKMSDLL 479
Qy 480 PMTMLPSIIFTCIVVFMGLKPKADAFVVMFTLMVAYSSSMALAIAGQSVSVATL 539
Db 480 PMRFLPSVIFTCILYFMLGKKTVDAFFIMFTLMVAYTASSMALALATGQSVSVATL 539
Qy 540 LMTICFVFMFIFSGLLVNLTTIASLWQYFSIPRYGFTALQHNFPNGFCPLNATG 599
Db 540 LMTIAFVFMFIFSGLLVNLRTIGPWSLWQYFSIPRYGFTALQHNFPNGFCGFNWD 599
Qy 600 NNPC--NYATCTGEEYLVKQIDLSPWGLWKNHVALACMIVIFLTAYLKLFLKKYS 655
Db 600 NSTCVNSYAICTGNEYLYNQIELSPWGLWKNHVALACMIIIFLTAYLKLFLKKYS 657

RESULT 8
Q9R0W57 PRELIMINARY; PRT; 657 AA.
ID Q9R0W57
AC Q9R0W57;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE ABC transporter ABCG2.
GN ABCG2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=wistar;
RA Hori S., Ohtsuki S., Terasaki T.;
RT "Expression and regulation of ABCG2 at the rat blood-brain barrier.";
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB108817; BAC76396.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR006162; Ppantne_S.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
DR PROSITE; PS00012; PHOSPHOPANTHEINE; 1.
SQ SEQUENCE 657 AA; 72960 MW; C975C61A08489027 CRC64;

Query Match      82.2%; Score 2754; DB 11; Length 657;
Best Local Similarity 81.0%; Pred. No. 1.7e-182;
Matches 533; Conservative 52; Mismatches 69; Indels 4; Gaps 3;

QY 1 MSSNVEVPIPVSGQNTGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRPVE 60
DB 1 MSSNDHVLVPMSQRKNGLPMSRGARTLAEGDVLSPFHITTYRVKVSGLF-VKTA 59

QY 61 KEILSNINGIMKPGNLAILGPTGGKSSLLDLAARKDPSGLSGDLVINGAPPAFKCN 120
DB 60 KEILSDINGIMKPGNLAILGPTGGKSSLLDLAARKDPSGLSGDLVINGAPPAFK 119

QY 121 SGYVQDDVVMGTLTVRENLOFSAALRLATTNTNHEKNERINRVIOELGLDKVADSKVGT 180
DB 120 SGYVQDDVVMGTLTVRENLOFSAALRLATTNTNHEKNERINTIILKELGLDKVADSKVGT 179

QY 181 QFIRGVSGERKRTSGMELITDPSILFLDEPTTGLDSDSTANAVLLLLKRMKSGQRTIIF 240
DB 180 QFTRGISGGERKRTSGMELITDPSILFLDEPTTGLDSDSTANAVLLLLKRMKSGQRTIIF 239

QY 241 SIHQPRYSIFKLPDSLTLLASGRMLPHGPAQALGFYFESAGYHCEAYNNPADFFLDVING 300
DB 240 SIHQPRYSIFKLPDSLTLLASGRMLPHGPAQALGFYFESAGYHCEAYNNPADFFLDVING 299

QY 301 DSTAVALNR-EEDFKATEIIEPSKQDKPIELKAEIYVNSFSFYKETKAEHLQSGGKKK 359
DB 300 DSSAVMLNRGEQDHEANKTEEPSKREKPIIENLAEFYINSTIYGETKAEHLQSGGKKK 359

QY 360 KITVFEKISYTTSCFCHLRWVSKSPKLLGNPOASIAQIIVTVVLGIVGAIYFGLKND 419
DB 360 GSSAFREPVVYTFCHQLRWIARRSFKNLLGNPOASVAQLIVTVLGLIIGALYFGLKND 419

QY 420 STGQNRAAGVLFLLTNNQCFSSYSAVELFVVEKKLFIHEYISGYRVSSYFGLKLSDLL 479
DB 420 PTGQNEAGVYFFLLTNNQCFSTVSARELVVEKKLFIHEYISGYRVSSYFGLKLSDLL 479

QY 480 PMTWLPSIIFTCIVYFVLMGLKPKADAFVMMFTLMVAYSASSMALAIAAGQSVSVATL 539
DB 480 PMRFLPSVITCLDYFVLMGLKRTVEAFFIMMFTLMVAYSASSMALAIAAGQSVSVATL 539

QY 540 LMTICFVPMIFSGLLVNLATIASLWSLQYFSIPRYGFTALQHNFFLQGNFCPLNATG 599
DB 540 LMTISFVFMVLFSGLLVNLRTIGPWSLWQYFSIPRYGFTALQHNFFLQGNFCPLNATG 599

QY 600 NNPC--NYATCTGVEYLVKQIDLSPMGLWKNHVALACMIVILTTAYLKLFLKKYS 655
DB 600 NSTCNSYTTCTGNDYLIQIDLSPMGLWKNHVALACMIIIFLTAYLKLFLKKYS 657

RESULT 9
Q80ST1 PRELIMINARY; PRT; 657 AA.
ID Q80ST1
AC Q80ST1;

```

```

DT 01-JUN-2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE ATP-binding cassette protein G2 transcript variant B (ATP-binding
DE cassette protein G2 transcript variant C) (ATP-binding cassette
DE protein G2 transcript variant A).
DE ABCG2.
GN Rattus norvegicus (Rat).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RA Yabuuchi H., Ishikawa T.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY089996; AAM09106.1; -.
DR EMBL; AY089997; AAM09107.1; -.
DR EMBL; AY089998; AAM09108.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR006162; Ppantne_S.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
DR PROSITE; PS00012; PHOSPHOPANTHEINE; 1.
KW ATP-binding.
SQ SEQUENCE 657 AA; 72960 MW; E194871E1C1AC201 CRC64;

Query Match      82.1%; Score 2752; DB 11; Length 657;
Best Local Similarity 80.9%; Pred. No. 2.3e-182;
Matches 532; Conservative 53; Mismatches 69; Indels 4; Gaps 3;

QY 1 MSSNVEVPIPVSGQNTGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRPVE 60
DB 1 MSSNDHVLVPMSQRKNGLPMSRGARTLAEGDVLSPFHITTYRVKVSGLF-VKTA 59

QY 61 KEILSNINGIMKPGNLAILGPTGGKSSLLDLAARKDPSGLSGDLVINGAPPAFKCN 120
DB 60 KEILSDINGIMKPGNLAILGPTGGKSSLLDLAARKDPSGLSGDLVINGAPPAFKCS 119

QY 121 SGYVQDDVVMGTLTVRENLOFSAALRLATTNTNHEKNERINRVIOELGLDKVADSKVGT 180
DB 120 SGYVQDDVVMGTLTVRENLOFSAALRLATTNTNHEKNERINTIILKELGLDKVADSKVGT 179

QY 181 QFIRGVSGERKRTSGMELITDPSILFLDEPTTGLDSDSTANAVLLLLKRMKSGQRTIIF 240
DB 180 QFTRGISGGERKRTSGMELITDPSILFLDEPTTGLDSDSTANAVLLLLKRMKSGQRTIIF 239

QY 241 SIHQPRYSIFKLPDSLTLLASGRMLPHGPAQALGFYFESAGYHCEAYNNPADFFLDVING 300
DB 240 SIHQPRYSIFKLPDSLTLLASGRMLPHGPAQALGFYFESAGYHCEAYNNPADFFLDVING 299

QY 301 DSTAVALNR-EEDFKATEIIEPSKQDKPIELKAEIYVNSFSFYKETKAEHLQSGGKKK 359
DB 300 DSSAVMLNRGEQDHEANKTEEPSKREKPIIENLAEFYINSTIYGETKAEHLQSGGKKK 359

QY 360 KITVFEKISYTTSCFCHLRWVSKSPKLLGNPOASIAQIIVTVVLGIVGAIYFGLKND 419
DB 360 GSSAFREPVVYTFCHQLRWIARRSFKNLLGNPOASVAQLIVTVLGLIIGALYFGLKND 419

QY 420 STGQNRAAGVLFLLTNNQCFSSYSAVELFVVEKKLFIHEYISGYRVSSYFGLKLSDLL 479
DB 420 PTGQNEAGVYFFLLTNNQCFSTVSARELVVEKKLFIHEYISGYRVSSYFGLKLSDLL 479

QY 480 PMTWLPSIIFTCIVYFVLMGLKPKADAFVMMFTLMVAYSASSMALAIAAGQSVSVATL 539
DB 480 PMRFLPSVITCLDYFVLMGLKRTVEAFFIMMFTLMVAYSASSMALAIAAGQSVSVATL 539

QY 540 LMTICFVPMIFSGLLVNLATIASLWSLQYFSIPRYGFTALQHNFFLQGNFCPLNATG 599
DB 540 LMTISFVFMVLFSGLLVNLRTIGPWSLWQYFSIPRYGFTALQHNFFLQGNFCPLNATG 599

QY 600 NNPC--NYATCTGVEYLVKQIDLSPMGLWKNHVALACMIVILTTAYLKLFLKKYS 655
DB 600 NSTCNSYTTCTGNDYLIQIDLSPMGLWKNHVALACMIIIFLTAYLKLFLKKYS 657

RESULT 9
Q80ST1 PRELIMINARY; PRT; 657 AA.
ID Q80ST1
AC Q80ST1;

```

```

Db 480 PMRFLPSVIYTCILYFMLGLKRTVEAFFIMFTLIWVAYTASSMALAIAAGQSVSVATL 539
QY 540 LMTICFVFMIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNELGQFCGLNATG 599
Db 540 LMTISFVFMIFSGLLVNLRTIGPWSLWLOYSIPRYGFTALQHNELGQFCGLNATG 599
QY 600 NNPC--NYATCTGEEYLVKQGDLSLPGWLNKHNVALACMIVIFLTAYLKLFLKKYS 655
Db 600 NSTCVNSYITCTGNDYLIHQIDLSLPGWLNKHNVALACMIIIFLTAYLKLFLKKYS 657

RESULT 10
Q80XF3 ID Q80XF3 PRELIMINARY; PRT; 657 AA.
AC Q80XF3;
DT 01-JUN-2003 (T-REMBLrel. 24, Created)
DT 01-JUN-2003 (T-REMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (T-REMBLrel. 25, Last annotation update)
DE ATP-binding, cassette transporter ABCG2.
GN ABCG2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Shimano K., Satake M., Okaya A., Kitanaka J., Kitanaka N.,
RA Takemura M., Sakagami M., Terada N., Tsujimura T.;
RT "Hepatic Oval Cells Have the Side Population Phenotype Defined by
RT Expression of ATP-binding Cassette Transporter ABCG2/BCRP1.";
RL Am. J. Pathol. 0:0-0(2003).
DR EMBL; AB094089; BAC75666.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.
DR GO; GO:000166; P:nucleotide binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR GO; GO:0003593; AAA ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR006162; Ppantne.S.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SW00382; AAA; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
DR PROSITE; PS00012; PHOSPHOTRANSFERASE; 1.
DR ATP-binding.
KW ATP-binding.
SQ SEQUENCE 657 AA; 72961 MW; 45980CC3903DSC5 CRC64;

Query Match 81.8%; Score 2742; DB 11; Length 657;
Best Local Similarity 80.9%; Pred. No. 1.1e-181;
Matches 532; Conservative 51; Mismatches 71; Indels 4; Gaps 3;

QY 1 MSSNVEVPIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPVE 60
Db 1 MSSNDHVLVPMQSRNKLPGMSRGARTLAEGDVLSPHITTVRVKVGFL-VRKTA 59
QY 61 KEILSNINGIMKPGNALIGTGGKSSLLDVLAAKDPGLSGDVLINGAPRANPKCN 120
Db 60 KEILSDINGIMKPGNALIGTGGKSSLLDVLAAKDPGLSGDVLINGAPRANPKCS 119
QY 121 SGVVVDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIOELGDKVADSKVGT 180
Db 120 SGVVVDDVVMGTLTVRENLOFSAALRLPFAKMKTHEKNERINTIIEKLGDKVADSKVGT 179
QY 181 PIRGVSGEKKRTSIGMELITDPSILFDEPTTGLDSSSTANAVLLILKMSKQRTIIF 240
Db 180 QFTRIGSGEKKRTSIGMELITDPSILFDEPTTGLDSSSTANAVLLILKMSKQRTIIF 239
QY 241 SIHQPRYSIFKLFDSLTLASGLMHPGPAQEAALGYFESAGHCEAYNNPADPFLDLING 300
Db 240 SIHQPRYSIFKLFDSLTLASGLMHPGPAQEAALGYFESAGHCEAYNNPADPFLDLING 299
QY 301 DSTAVALNR-BEDFKATEIIEPSKQDKPLIEKLAIEIYVNSGPFYKETAELHQLSGEKKK 359

```

```

Db 300 DSSAVMLNRGEQHEANKTEEPSKREKPIENLAEFYINSTIYGETKAELOQLPVAQKKK 359
QY 360 KITVKEISVTTTFCQHLRWVSKRSFKNLGNPQASIAQIIVTVLGLVIGATYFGLKND 419
Db 360 GSDFPEPVYTFCHQLRWIARSRFKNLGNPQASVAQIIVTVLGLIIGALYFGLKND 419
QY 420 STGIQNRAGVFLFTTNQCPSSVASVELFVVEKLLFIHEYISGYRVSSYFGLKLSDLL 479
Db 420 PTGMQNRAGVFLFTTNQCPFTSVAVELFVVEKLLFIHEYISGYRVSSYFGLKLSDLL 479
QY 480 PMTMLPSIIFTCTIVYFMLGLKPKADAFFVNMFTLMVAVSASSMALAIAAGQSVSVATL 539
Db 480 PMRFLPSVIYTCILYFMLGLKRLVEAFFIMFTLIWVAYTASSMALAIAAGQSVSVATL 539
QY 540 LMTICFVFMIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNELGQFCGLNATG 599
Db 540 LMTISFVFMIFSGLLVNLRTIGPWSLWLOYSIPRYGFTALQHNELGQFCGLNATG 599
QY 600 NNPC--NYATCTGEEYLVKQGDLSLPGWLNKHNVALACMIVIFLTAYLKLFLKKYS 655
Db 600 NSTCVNSYITCTGNDYLIHQIDLSLPGWLNKHNVALACMIIIFLTAYLKLFLKKYS 657

RESULT 11
Q8BK15 ID Q8BK15 PRELIMINARY; PRT; 650 AA.
AC Q8BK15;
DT 01-MAR-2003 (T-REMBLrel. 23, Created)
DT 01-MAR-2003 (T-REMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (T-REMBLrel. 25, Last annotation update)
DE ATP-binding cassette.
GN ABCG3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=Eye;
RC MEDLINE=22354683; PubMed=12466851;
RX The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RL Nature 420:563-573(2002).
DR EMBL; AK051880; BAC34799.1; -.
DR FIR; PT0645; PT0645.
DR MGD; MGI:1351624; Abcg3.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
DR PROSITE; PS00012; PHOSPHOTRANSFERASE; 1.
DR ATP-binding.
SQ SEQUENCE 650 AA; 73613 MW; B9395BDBDFD945E1 CRC64;

Query Match 53.3%; Score 1787.5; DB 11; Length 650;
Best Local Similarity 55.3%; Pred. No. 1.8e-115;
Matches 363; Conservative 104; Mismatches 174; Indels 15; Gaps 5;

QY 1 MSSNVEVPIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPVE 60
Db 1 MASNNDPTVISMIERHLCDLPTNTSDTLTLEZAVLSFHNISYQFTVQSGFPLKCAV 60
QY 61 KEILSNINGIMKPGNALIGTGGKSSLLDVLAAKDPGLSGDVLINGAPRANPKCN 120
Db 61 IERLSNISGIMKPGNALIMGPQDGRSLLDVLAAARDPGLSGDILINGKRPANFKCT 120
QY 121 SGVVVDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIOELGDKVADSKVGT 180
Db 121 SGVVPQNDVVMGTLTVRDNIETFSALRLPVTITRDEKRRINEVLELLHLNK-----E 173

```



RT "Full length cDNA of gene At5g06530 (GI:15240083).";  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,  
 RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,  
 RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,  
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
 RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,  
 RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,  
 RA Theologis A.;  
 RT "Arabidopsis Open Reading Frame (ORF) Clones.";  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 RA EMBL: AY059787; AL24135.1; --  
 RA EMBL: AY114068; AN45116.1; --  
 DR GO: 0016020; C:membrane; IEA.  
 DR GO: 0005524; P:ATP binding; IEA.  
 DR GO: 0004009; P:ATP-binding cassette (ABC) transporter acti. .; IEA.  
 DR GO: 0006810; P:transport; IEA.  
 DR InterPro: IPR003439; ABC transporter.  
 DR Pfam: PF00005; ABC tran; 1.  
 DR ProDom: PD000006; ABC transporter; 1.  
 DR PROSITE: PS00211; ABC\_TRANSPORTER\_1; 1.  
 DR PROSITE: PS00893; ABC\_TRANSPORTER\_2; 1.  
 DR PROSITE: PS00042; PHOSPHOTANTHEINE; 1.  
 SQ SEQUENCE 751 AA; 82931 MW; 5564454F99D0A2EC CRC64;  
 Query Match 24.1%; Score 809; DB 10; Length 751;  
 Best Local Similarity 33.9%; Pred. No. 1.7e-47;  
 Matches 216; Conservative 110; Mismatches 245; Indels 66; Gaps 16;  
 37 LSFNICYRVLKSGFLPCRKPKVEKILSNINGIMKPG-INAILGPTGGKSKSLLDVLA 95  
 157 LKFDVTVKVIK----LTSSVEKITGTSGVNPGEALMGPSGSKTLLSLLAG 212  
 96 RKDPSGLSGDLINGAPRANFKCNVGVQDDVVMGTLTVRENLOFSAALRLATMTNH 155  
 213 RISQSTGSGVTVNDKPKYKYLKSGIFGVTDVLFPHLTIVKTYVAELRLPKLTRE 272  
 156 EKBNRINRVIOELGLDVKADSVKQTFIRGVSGGERKRTISGMELITDPSILFDPTTG 215  
 273 QKQRLDIVDELGLERQDQTMIGAFVGVSGGERKRVISGNBIINPSSLILDEPTSG 332  
 216 LDSSTANAVLLKXMSKQGTIIFTSHQPRYSIFKLFDSITLLASGLMFGHPAQBALG 275  
 333 LDSTALTILMLHDAEAGKTVITTHQPSRULFHRFDKILLGRSLYFGKSEALD 332  
 276 YFESAGYHCEAYNPNADPFLDIINGDSTAVALNREEDFKA-----TEIIPSKQDKP 327  
 393 YFSSIGCSPLIAMPNPAEFLLLDANGNINDISVPSELDRVQVGNSGRETQTKPS--PAA 450  
 328 LIEKLAELIYVNSSFYKTEKALHOLS-GGEKKKITVFE-----ISYTSFCHOLRW 379  
 451 VHEYLVEAYETRVABQEKLLDPVLDDEAKAKSTRIKROWGTCWEQYCIIFCRGLKE 510  
 380 VSKRSKNLGNPQASIAQIITVVVLGLVIGAIYFGLKND---STGIQNRAGVLFPLTTN 436  
 511 RRHEVFSWL-----RVTVQLSTAV---ILGLLW--QSDIRTPWGLDQAGLFFIAVE 559  
 437 QCFSSV-SAVELFVVEKKLFTHEYISGYRVSSFLGKLLSDILPMTLPSIIITCTVYF 495  
 560 WGFVFVFTAFAPQERAMLNKERAADMYRLSAYPLARTTSDLPDLFLPLS-LFLLVYVF 618  
 496 MLGLPKADAPVMMFTLMVAYSASSMALAIAGQSVSVVATILMTICFVFMIFSGLL 555  
 619 MTGLRISYPPFLSMLTVFLCIIRAQGLGAIGALIMDLKATTLASVTVMTFLAGGFF 678  
 556 VNLATTIAGWLSQVFSIPRYGFTALQHNEFLQGNFCGLNATGNPNYATCTGBEYIV 615  
 679<sub>4</sub>VKKVPV--FISWIRYLSFNHYTKLLKLVQY--QDFAVNSING-----MRI 719

Qy 616 KQIDLSPLWGLKRWKHWALACMIVIFITIAVLKLLFLK 652  
 Db 720 DNGL-----TEVAALVVMIFGYRLLAYLSRQMK 748

RESULT 14  
 Q9NH94 PRELIMINARY; PRT; 687 AA.  
 AC Q9NH94;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE ATP dependent transmembrane transporter protein.  
 GN WH3.  
 OS Bombyx mori (Silk moth).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea;  
 OC Bombycidae; Bombyx.  
 OX NCBI\_TaxID=7091;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=kin-Shiu X Sho-wa;  
 RA MEDLINE=20469043; PubMed=11016828;  
 RX Abraham E.G., Sezutsu H., Kanda T., Sugasaki T., Shimada T.,  
 RA Tamura T.;  
 RT "Identification and characterization of a silkworm ABC transporter  
 RT gene homologous to Drosophila white.";  
 RT Mol. Gen. Genet. 264:11-19(2000).  
 CC -|- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.  
 DR EMBL: AF229609; AAF61569.1; --  
 DR GO: 0016021; C:integral to membrane; IEA.  
 DR GO: 0005524; P:ATP binding; IEA.  
 DR GO: 0004009; P:ATP-binding cassette (ABC) transporter acti. .; IEA.  
 DR GO: 0000166; P:nucleotide binding; IEA.  
 DR GO: 0006810; P:transport; IEA.  
 DR InterPro: IPR003593; AAA\_ATPase.  
 DR InterPro: IPR003439; ABC transporter.  
 DR Pfam: PF00005; ABC tran; 1.  
 DR ProDom: PD000006; ABC transporter; 1.  
 DR SMART: SM00382; AAA; 1.  
 DR TIGRFAMs: TIGR00955; 3a01204; 1.  
 DR PROSITE: PS00211; ABC\_TRANSPORTER\_1; 1.  
 DR PROSITE: PS00893; ABC\_TRANSPORTER\_2; 1.  
 DR ATP-binding; Transmembrane; Transport.  
 KW SEQUENCE 687 AA; 75835 MW; ECD336333F0991AB CRC64;  
 Query Match 24.1%; Score 808; DB 5; Length 687;  
 Best Local Similarity 32.5%; Pred. No. 1.7e-47;  
 Matches 200; Conservative 108; Mismatches 248; Indels 60; Gaps 12;  
 61 KEILNINIMKPG-INAILGPTGGKSKSLLDVLAARKDPSGL--SGDVLINGAP-RPAN 116  
 100 KQLLRNVGAAYFGEALLAIMGSSGAGKTTLLTLTFR-TPGVVAITGTRALNGQATPDA 158  
 117 FKNCSYVVDVVMGTLTVRENLOFSAALRLATMTNHKNERINRVIOELGLDKVADS 176  
 159 LIALSAVVOQDLFGTLTVREHLVFAQVMDRIHPYAQRMKRVQEVIELALSKCQNT 218  
 177 KVGTQ-FIRGVSGGERKRTISGMELITDPSILFDPTTGSDSTANAVLLKXMSKQ 235  
 219 VIGIPGRKXISGGEMKRLSPASEVLTPPLMFCDEPTSGDLSFMAQNVIOVLKGLAQK 278  
 236 RTIIFSIHQPRYSIFKLFDSITLLASGLMFGHPAQEALGYFESAGYHCEAYNPNADPFL 295  
 279 KTVCTVTHQPSSELVAMFDKLLIMADGRVAFVFGSSDEAFQFKELGACFANYNPADHFI 338  
 296 DIINGDSTAVLNREEDFKATIEIPSKQDKPLIEKLAELIYVNSSFYKTEKALHOLS 355  
 339 QLLAG-----VPOREEVTRHT-----IDTVCTAFKSEIGCRFAAEAEENALYN 381  
 356 EKK-----KKITVFEISYTSFCHOLRWVSKSPKLLGNPQASIAQIITVV 403

Db 382 ERKICAGLADAPWANSSTTRAGSPYKASWCTQFRAVLWRSLWTKPEMLIKVFLQTI 441

Qy 404 VLGLVIGAIYFGLKNDSTGIONRAGVLFLLTNQCFSSVSVA-ELFVVEKKLFIHEVISG 462

Db 442 MYSILIGVYFQNLDDQGVNKGAIFFMLTNMTFQNFIAVINVCSELPIFIREHSG 501

Qy 463 YVRVSSYFGLKLLDLEPMTWLPSTIIFCIVVFMGLKPKADAFVMMFTLMVAYGASS 522

Db 502 MYRADVYFSLKTLAEPVFAIP-LVFTTIAYMGLNPDKPRFFIASGLAALVTVNATS 560

Qy 523 MALAJAAGQSVSVATLMTICFVMMIFPSGLLVNLTITIASWLSWLOYSFIPRGFTALQ 582

Db 561 FGYLLSCASSSVMAASGPPIIIFMLFGFFLNSGVPYLSWISYLSMFHYGNALL 620

Qy 583 HNEFLG-----CNFCPLNATGNNPCNVATCTGEEYLVKQGDLSPLWGLKXHVALA 634

Db 621 INQWAGVETIATRENF-----TCPASQGVLEILSQCDDFAMDVVNMI 665

Qy 635 CMIVIFLTIAYLKLLF 650

Db 666 LLFVGRFRLAYLALLW 681

RESULT 15

Q94960 PRELIMINARY; PRT; 687 AA.

AC Q94960;

DT 01-FEB-1997 (TrEMBLrel. 02, Created)

DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE White.

GN W OR WHITE.

OS Drosophila simulans (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7240;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=SIM;

RX MEDLINE=96170784; PubMed=8601488;

RA Kirby D.A., Stephan W.;

RT "Haplotype test reveals departure from neutrality in a segment of the white gene of Drosophila melanogaster.";

RL Genetics 141:1483-1490(1995).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=SIM;

RA Kirby D.A., Stephan W.;

RL Genetics 0:0-0(0).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=SIM;

RA Kirby D.A., Stephan W.;

RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.

CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.

DR EMBL; U64875; AAB06578.1; -

DR FlyBase; FBgn0012902; Dsim.w.

DR GO; GO:0016020; C:membrane; IEA.

DR GO; GO:0005524; F:ATP binding; IEA.

DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti...; IEA.

DR GO; GO:0000166; F:nucleotide binding; IEA.

DR GO; GO:0006810; P:transport; IEA.

DR InterPro; IPR003593; AAA\_Atpase.

DR InterPro; IPR005284; Pigment\_permease.

DR Pfam; PF00005; ABC\_tran; 1.

DR ProDom; PD000006; ABC\_transporter; 1.

DR SMART; SMO0382; AAA; 1.

DR TIGRFAMs; TIGR00955; 3a01204; 1.

DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.

DR PROSITE; PS00893; ABC\_TRANSPORTER\_2; 1.

KW ATP-binding; Transport.

SQ SEQUENCE 687 AA; 75721 NW; D02DFEE299A25592 CRC64;

Query Match 24.0%; Score 806; DB 5; Length 687;

Best local Similarity 32.1%; Pred. No. 2.4e-47;

Matches 210; Conservative 133; Mismatches 252; Indels 60; Gaps 17;

Qy 5 NVEVPIPSQNGTNGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCKRVEKIL 64

Db 74 NMDIFGAVNQ-----PGSGWRQLVNRIRGLFCNERHI-----PAPR---KHLL 113

Qy 65 SNINGIMKGG-INAILGPTGGKSSILDVLAARKDPSGL-----SGDVILINGAPRA-NFK 118

Db 114 KNCVGVAYPEGLAVLWVGSGAGKTTLLNALAFR-SFGIQVSPSGWRLLNGPVDAKEMQ 172

Qy 119 CNSGYVQDDVVMGTLTVRENLQFSAALRLATMTNHEKNERINRVIQELGLDKVADSKV 178

Db 173 ARCAVYQDDLLFIGSLTAREHLIFQAMMRPHLYRQVRVARVDQVIELLSKQHTII 232

Qy 179 GTO-FIRGVSGGERKRTSIGMELITDPSITLFDDEPTGLDSTANAVLLLLKRMKQGR 237

Db 233 GVGGRAGLSGGERKRLAFASEALTDPPLIICDETSGLDSTAHVVQVLKLSQKGT 292

Qy 238 IIPSIHQPRYSIFKLFDSTLLASGRLMFHPGAQOALGYFESAGYHCEAYNNPAPFFDI 297

Db 293 VILTIHQPSSELFELFDKILLMAEGKRVAFGLTPSEAVDPFSSVGAQCTPNYPADFYQV 352

Qy 298 INGDSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIVVNSFYKETKAEHLQSGGK 357

Db 353 L-----AVVPGREIESR-----DRIAKICDNFAISKVAR-DMEQLLATQN 391

Qy 358 KKKITVKEISYT--TSPCHQLRWYSKRSFKNLLGNPQASIAQIIVTVVLGIVIGAIYFG 415

Db 392 LEKPLEQPENGITYKATWFMQFRAVLWRSLWLSVLEKFLVVKVLIQTTWVAILGLIFLG 451

Qy 416 LKNDSTGIONRAGVLFLLTNQCFSSVA-VELFVVEKKLFIHEVISGYRVVSSYELGKL 474

Db 452 QQLTQVGMNNGAIFLFLTNMTFQNVFATINVTSELFPVREARSRLYRCDTYFLGKT 511

Qy 475 LSDLPLMTLPSTIIFCIVVFMGLKPKADAFVMMFTLMVAYGASSMALAIAAGQSVV 534

Db 512 IAE-LPLFLTVPLVFTAIAYPMGLRAGVMHFNCLALVTLVANVSTSFGLISCASSST 570

Qy 535 SVATLLMTICFVPMIFPSGLLVNLTITIASWLSWLOYSFIPRGFTALQHNELFQNFQFCG 594

Db 571 SMALSVGPVVIIPFLFGGFFLNSGSPVYLVKLSYLSWFRYANEGLLINQWADVE--FG 628

Qy 595 -LNATGNPCNVATCTGEEYLVKQGDLSPLWGLKXHVALA CMIVIFLTIAYLK 648

Db 629 EISCTSN-----TTCFSSGKVILETLNFAADLPDYVGLAILIVSFQVLAYLAL 679

Search completed: March 23, 2004, 18:31:03

Job time : 49 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2004, 18:23:53 ; Search time 18 seconds  
(without alignments)  
1894.775 Million cell updates/sec

Title: US-09-961-086-1  
Perfect score: 3352  
Sequence: 1 MSSSNVEFIPVSGQTNGF.....MIVIFLTIAVLKLLFLKKYS 655

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3339	99.6	655	1 ABG2 HUMAN	Q9UNQ0 homo sapien
2	1786.5	53.3	650	1 ABG3 MOUSE	Q99P81 mus musculus
3	835.5	24.9	1049	1 ADP1 YEAST	P25371 saccharomyc
4	812	24.2	687	1 WHIT DROME	P10090 drosophila
5	772	23.0	679	1 WHIT CERCA	Q17320 ceratidis c
6	754.5	22.5	677	1 WHIT LUCCU	Q05360 lucilia cup
7	751	22.4	695	1 WHIT ANOGA	Q27256 anopheles g
8	736.5	22.0	709	1 WHIT ANOAL	Q16928 anopheles a
9	714.5	21.3	666	1 ABG1 MOUSE	C64343 mus musculus
10	708	21.1	678	1 ABG4 HUMAN	Q9H172 homo sapien
11	706.5	21.1	678	1 ABG5 HUMAN	P45844 homo sapien
12	694.5	20.7	652	1 ABG6 MOUSE	Q99P88 mus musculus
13	686.5	20.5	598	1 YPC3 CAHEL	Q11180 caenorhabdi
14	682.5	20.4	651	1 ABG5 HUMAN	Q9H222 homo sapien
15	682.5	20.4	652	1 ABG5 RAT	Q99P87 rattus norv
16	669.5	20.0	694	1 ABG8 RAT	P58428 rattus norv
17	665	19.8	673	1 ABG8 MOUSE	Q9DBM0 mus musculus
18	658.5	19.6	1294	1 YOH5 YEAST	Q08234 saccharomyc
19	654.5	19.5	666	1 SCRT DROME	P45843 drosophila
20	639.5	19.1	673	1 ABG8 HUMAN	Q9H221 homo sapien
21	627.5	18.7	610	1 YQ5C CAHEL	Q09466 caenorhabdi
22	612.5	18.3	1501	1 SNQ2 YEAST	P32568 saccharomyc
23	612	18.3	1530	1 BFRI SCHPO	P41820 schizosacch
24	602	18.0	1511	1 PDRE YEAST	P33302 saccharomyc
25	594	17.7	1564	1 PDRE CANAL	P51533 saccharomyc
26	593.5	17.7	1499	1 YN99 YEAST	P78595 candida alb
27	585	17.5	1333	1 YN99 YEAST	P53756 saccharomyc
28	580.5	17.3	1333	1 PDRE YEAST	Q04182 saccharomyc
29	544	16.2	1511	1 PDRE YEAST	Q02785 saccharomyc
30	529	15.8	1501	1 CDR3 CANAL	O42690 candida alb
31	528	15.8	1501	1 CDR1 CANAL	P43071 candida alb
32	500	14.9	1490	1 CDR4 CANAL	O74676 candida alb
33	488.5	14.6	1410	1 PDRE YEAST	P40550 saccharomyc

P12428 drosophila  
Q24739 drosophila  
Q95342 homo sapien  
P74548 synchocyst  
Q82080 anabaena sp  
Q99758 homo sapien  
P46903 bacillus su  
P02914 escherichia  
C28882 archaeglob  
P22638 anabaena sp  
Q00449 drosophila  
Q82W5 nitrosomona

34 474 14.1 675 1 BROW DROME  
35 455 13.6 668 1 BROW DROVI  
36 243 7.2 1321 1 AB11 HUMAN  
37 242 7.2 355 1 CYSA-SNV3  
38 240 7.2 338 1 CYSA-ANASP  
39 238.5 7.1 1704 1 ABG3 HUMAN  
40 238 7.1 246 1 NATA-BACSU  
41 235 7.0 371 1 MALK-ECOLI  
42 234.5 7.0 236 1 LIVF-ARCFU  
43 229.5 6.8 607 1 HEPA-ANASP  
44 229 6.8 1302 1 MDR4-DROME  
45 228.5 6.8 362 1 CYSA-NITEU

ALIGNMENTS

RESULT 1  
ABG2 HUMAN STANDARD; PRT; 655 AA.  
AC Q9UNQ0; Q95374; Q9BY73; Q9NUS0;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE ATP-binding cassette, sub-family G, member 2 (Placenta-specific ATP-binding cassette transporter) (Breast cancer resistance protein).  
DE binding cassette transporter OR BCRP OR BCRP1.  
GN ABG2 OR ABG2 OR BCRP OR BCRP1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=99065313; PubMed=9850061;  
RA Allikmets R., Schriml L.M., Hutchinson A., Romano-Spica V., Dean M.;  
RT "A human placenta-specific ATP-binding cassette gene (ABCP) on chromosome 4q22 that is involved in multidrug resistance.";  
RL Cancer Res. 58:5337-5339(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Breast cancer;  
RX MEDLINE=99080071; PubMed=9861027;  
RA Doyle L.A., Yang W., Abruzzo L.V., Krogmann T., Gao Y., Rishi A.K., Ross D.D.;  
RT "A multidrug resistance transporter from human MCF-7 breast cancer cells.";  
RL Proc. Natl. Acad. Sci. U.S.A. 95:15665-15670(1998).  
RN [3]  
RP ERRATUM.  
RA Doyle L.A., Yang W., Abruzzo L.V., Krogmann T., Gao Y., Rishi A.K., Ross D.D.;  
RT "Breast cancer resistance protein constitutes a 140-kDa complex as a homodimer.";  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE OF 198-655 FROM N.A.  
RC TISSUE=Placenta;  
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y., Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M., Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S., Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K., Yamanoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Ninomiya K., Iwayanagi T.;  
RT "NEO human cDNA sequencing project.";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
RN [6]

RP REVIEW.  
 RX MEDLINE=21474438; PubMed=11590207;  
 RA Schmitz G., Langmann T., Heimerl S.;  
 RT "Role of ABCG1 and other ABCG family members in lipid metabolism.";  
 RL J. Lipid Res. 42:1513-1520(2001).  
 CC !- FUNCTION: Xenobiotic transporter that appears to play a major role  
 CC in the multidrug resistance phenotype of a specific MCF-7 breast  
 CC cancer cell line. When overexpressed, the transfected cells become  
 CC resistant to mitoxantrone, daunorubicin and doxorubicin, display  
 CC diminished intracellular accumulation of daunorubicin, and  
 CC manifest an ATP-dependent increase in the efflux of rhodamine 123.  
 CC !- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
 CC !- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)  
 CC subfamily.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AF103796; AAC09188.1; -;  
 CC EMBL; AF098951; AAC97367.1; -;  
 CC EMBL; AB056867; BAB39212.1; -;  
 CC EMBL; AK002040; BAA92050.1; -;  
 CC Genew; HGNC:74; ABCG2.  
 CC MIM; 603756; -;  
 CC GO; GO:0016021; C: integral to membrane; TAS.  
 CC GO; GO:0005524; P: ATP binding; TAS.  
 CC GO; GO:0004009; P: ATP-binding cassette (ABC) transporter acti. . .; TAS.  
 CC GO; GO:0005215; P: transporter activity; TAS.  
 CC GO; GO:0008559; P: xenobiotic-transporting ATPase activity; TAS.  
 CC GO; GO:0009315; P: drug resistance; TAS.  
 CC GO; GO:0006810; P: transport; TAS.  
 CC InterPro; IPR003593; AAA ATPase.  
 CC InterPro; IPR003439; ABC transporter.  
 CC Pfam; PF00005; ABC\_tran; 1.  
 CC Pfam; PD000006; ABC\_transporter; 1.  
 CC SMART; SM00382; AAA; 1.  
 CC PROSITE; PS00211; ABC\_TRANSPORTER\_1; FALSE\_NEG.  
 CC PROSITE; PS00893; ABC\_TRANSPORTER\_2; 1.  
 CC ATP-binding; Transmembrane; Transport.  
 KW DOMAIN 1 395 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 396 416 POTENTIAL.  
 FT DOMAIN 417 428 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 429 449 POTENTIAL.  
 FT DOMAIN 450 477 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 478 498 POTENTIAL.  
 FT DOMAIN 499 506 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 507 527 POTENTIAL.  
 FT DOMAIN 528 535 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 536 556 POTENTIAL.  
 FT DOMAIN 557 630 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 631 651 POTENTIAL.  
 FT DOMAIN 652 655 CYTOPLASMIC (POTENTIAL).  
 FT NP\_BIND 80 87 ATP (POTENTIAL).  
 FT CARBOHYD 418 418 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 557 557 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 596 596 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 24 24 V -> A (IN REF. 2 AND 4).  
 FT CONFLICT 166 166 E -> Q (IN REF. 1).  
 FT CONFLICT 208 208 F -> S (IN REF. 1).  
 FT CONFLICT 315 316 MISSING (IN REF. 5).  
 FT CONFLICT 482 482 R -> T (IN REF. 2).  
 SQ SEQUENCE 655 AA; 72343 MW; 89A6D351DC5CCE0 CRC64;  
 Query Match 99.6%; Score 3339; DB 1; Length 655;  
 Best Local Similarity 99.5%; Pred. No. 66-222;  
 Matches 652; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MSSNNVEVFPVSGQNTNGFPATVSNDLKAFTEGAVLSFHNICVRVKLSGFLPCRKPE 60

Db 1 MSSNNVEVFPVSGQNTNGFPATVSNDLKAFTEGAVLSFHNICVRVKLSGFLPCRKPE 60  
 QY 61 KEILSNINGIMKPGINAILGPTGGKSSLLDVLAAKDPGSLGSDVLINGAPRANFKCN 120  
 Db 61 KEILSNINGIMKPGINAILGPTGGKSSLLDVLAAKDPGSLGSDVLINGAPRANFKCN 120  
 QY 121 SGYVQDDVVGTTITVRENLOFSAALRLATTTMTHKERNINRVIQELGLDKVADSKVGT 180  
 Db 121 SGYVQDDVVGTTITVRENLOFSAALRLATTTMTHKERNINRVIQELGLDKVADSKVGT 180  
 QY 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSSSTANAVLLLLKRMKSGKRIIF 240  
 Db 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSSSTANAVLLLLKRMKSGKRIIF 240  
 QY 241 SIHQPRYSIFKLFPSLTLLASGRIMFEGPACALGYFESAGYHCEAYNNPADFFLDIING 300  
 Db 241 SIHQPRYSIFKLFPSLTLLASGRIMFEGPACALGYFESAGYHCEAYNNPADFFLDIING 300  
 QY 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLABIYVNSFYKTKAELHOLSGGKKKK 360  
 Db 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLABIYVNSFYKTKAELHOLSGGKKKK 360  
 QY 361 ITVFKEISYTTSFCHQLRWKRSKSPKLLGNPQASIAQIIIVTVLGLVIGAIYFGLKND 420  
 Db 361 ITVFKEISYTTSFCHQLRWKRSKSPKLLGNPQASIAQIIIVTVLGLVIGAIYFGLKND 420  
 QY 421 TGIQNRAGVLPFLTTNQCFSSVSAVEFVVEKGLFIHEYISGYRVSVSYFGLKLSDLLP 480  
 Db 421 TGIQNRAGVLPFLTTNQCFSSVSAVEFVVEKGLFIHEYISGYRVSVSYFGLKLSDLLP 480  
 QY 481 MTWLPISIFTCIVYFMLGLKPKADAFVWMFTLMWVAYSASSMALAIAGQSVSVATLL 540  
 Db 481 MRMLPSIIFTCTIVYFMLGLKPKADAFVWMFTLMWVAYSASSMALAIAGQSVSVATLL 540  
 QY 541 MTICFVFMVIFSGLLVNLTTIASWLSWLYQFSIPRYGTALQHNFLGNCPCPLNATCN 600  
 Db 541 MTICFVFMVIFSGLLVNLTTIASWLSWLYQFSIPRYGTALQHNFLGNCPCPLNATCN 600  
 QY 601 NPCNVATCTGEEYLVKQIGIDLSFWGLWKNHVALACMIVIFLTIAVLLFLKKYS 655  
 Db 601 NPCNVATCTGEEYLVKQIGIDLSFWGLWKNHVALACMIVIFLTIAVLLFLKKYS 655  
 RESULT 2  
 ABG3\_MOUSE STANDARD; PRT; 650 AA..  
 ID ABG3\_MOUSE  
 AC Q99P81;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE ATP-binding cassette, sub-family G, member 3.  
 GN ABCG3.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 CX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Splesen;  
 RX MEDLINE=21030753; PubMed=11178751;  
 RA Mickleth L., Jain P., Miyake K., Schriml L.M., Rao K., Pojo T.,  
 RA Bates S., Dean M.;  
 RT "An ATP-binding cassette gene (ABCG3) closely related to the multidrug  
 RT transporter ABCG2 (MXR/ABCP) has an unusual ATP-binding domain.";  
 RL Mamm. Genome 12:86-88(2001).  
 CC !- SUBUNIT: May dimerize with another subunit to form a functional  
 CC transporter.  
 CC !- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
 CC !- TISSUE SPECIFICITY: Highest levels of expression in thymus and  
 CC spleen. Detected in lung and small intestine.  
 CC !- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)  
 CC subfamily.

CC	-1- CAUTION: Seems to have a defective ATP-binding region.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>	
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
CC	-----	
DR	EMBL; AF324242; AK14241.1; -	
DR	MGI; 1351624; Abcg3.	
DR	InterPro; IPR003439; ABC transporter.	
DR	Pfam; PF00005; ABC_tran; 1.	
DR	PROSITE; PS00211; ABC_TRANSPORTER_1; FALSE_NEG.	
DR	PROSITE; PS00893; ABC_TRANSPORTER_2; 1.	
KW	Transmembrane; Transport.	
FT	DOMAIN 1 387	
FT	TRANSMEM 388 408	
FT	DOMAIN 409 420	
FT	TRANSMEM 421 441	
FT	DOMAIN 442 469	
FT	TRANSMEM 470 490	
FT	DOMAIN 491 498	
FT	TRANSMEM 499 519	
FT	DOMAIN 520 527	
FT	TRANSMEM 528 548	
FT	DOMAIN 549 623	
FT	TRANSMEM 624 644	
FT	DOMAIN 645 648	
FT	SEQUENCE 650 AA; 73623 MW; 86A5ABE4DD26971C CRC64;	
Qy	Query Match 55.3%; Score 1786.5; DB 1; Length 650;	
Db	Best Local Similarity 53.3%; Pred. No. 3.2e-115;	
Db	Matches 363; Conservative 104; Mismatches 174; Indels 15; Gaps 5;	
Qy	1 MSSNVEFIPVSGQNGTPATASNDLKAPTEGAVLSFHNICVYVKLKGSLPCPKPVE 60	
Db	1 MASNNDPVTSMIRHEHCDLPETNTSBLKTUTEAVLSFHNISVQETVQSGFPLRKAYV 60	
Qy	61 KEILSNINGIMKGNAILGTGGKSLDVLAAKDPGSLGSDVLINGAPNPKCN 120	
Db	61 IERLSNIGIMKGNALMGQDGRSLLDVLAAKDPGSLGSDVLINGKPRANPKCT 120	
Qy	121 SGYVQDDVNGTITVRENLOFSAALRLATMTMHEKERNINRVIQELGDKVADSKVGT 180	
Db	121 SGYVPQNDVLTGTVDNLEFSAALRLPVTITRDEKRRRINEVLELHLNK-----E 173	
Qy	181 QPIRVSGGERKRTSIGMELTDPISILFDEPTTGLDSSSTANAVLLLLKRMKSGQRTIIF 240	
Db	174 QNIKPSKELKRRSIAMELVTEHPILFDDPTTGLDRLRTTDLVILVRMSKKGRTIIF 233	
Qy	241 SIHQPRYSIFKLFQSLTLLASRLMHFGPAQEAALGYFESAGYHCEAVNNPADPFLDLING 300	
Db	234 SINQPOYSIFKFFQSLTIVASGVKVMFHGPAQDALEYFRASAGYNESHNNPADPFLDVING 293	
Qy	301 DSTAVALNRE---EDFKATEIIEPSKODKPIELKLAELIYVNSVFKETKAEHLQSGEK 357	
Db	294 GFSNILDTEEDGHEDDKKEELFERQY---VTGKLANNYAQSPLYSETRALDQLLGEQK 350	
Qy	358 KKKITVFKIEISYTTSFCHQLRWVSRKSPKNLLGNPQASIAQIIVTVVLGLVIGAIYGLK 417	
Db	351 LERSAV-ETTCVTPFCQKQKWIICQSPKNKGFQFWTVIQAIITVILATVAGTAFRYLK 409	
Qy	418 NDSTGIONRAGVLPFLTNQCFSSYSAVELFVVEKKLFTHVEISGYYSVSSYFELKLLSD 477	
Db	410 NDCIEVQMRAGLLYLTTFQCITSVSAGELFVIDRVFRHHTSGYFVSSYFFGKLLAE 469	
Qy	478 LLEPMTLPSLIITCTIVYFNLGLKPKADAFVMMFTLMVAYSASMALAIAGOSVSVSA 537	
Db	470 LIPRLLPSTVSLITYVIAGVGMKMKCFPTMCTIMVLAYSASSPLSLIGAGENAVAP 529	
Qy	538 TLLMTICFVFMIFSGLLVNLTTIASWLSWLOQYFSIPRYGFTALQHNEFLQNGFCPLNA 597	

Db	530 TLLVTIYFVFWLPSGLSLYPSGFLPKLSIQYFSPHYGFRALLHNEFLGQNFCEHNT 589
Qy	598 TGNNPC-NYATCTGEEVLYVKQIDLSPLGWLKXKHNVALACMIVFLTIAYLKLFLK 652
Db	590 EVERSCHNYVICTGEEFLMIQGDLSWGFWENHLAVCTMILLITTYVQLLOVK 645

RESULT 3

ID	ADPI_YEAST	STANDARD;	PRT;	1049 AA.
AC	P25371;			
DT	01-MAY-1992 (Rel. 22, Created)			
DT	01-MAY-1992 (Rel. 22, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Probable ATP-dependent permease precursor.			
GN	ADPI OR YCR011C OR YCR11C OR YCR105.			
OS	Saccharomyces cerevisiae (Baker's Yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
OX	NCBI_TaxID=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92160395; PubMed=1789009;			
RA	Purnelle B., Skala J., Goffeau A.;			
RT	"The product of the YCR105 gene located on the chromosome III from			
RT	Saccharomyces cerevisiae presents homologies to ATP-dependent			
RT	permeases.";			
RL	Yeast 7:867-872(1991).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92327849; PubMed=1626432;			
RA	Skala J., Purnelle B., Goffeau A.;			
RT	"The complete sequence of a 10.8 kb segment distal of SUP2 on the			
RT	right arm of chromosome III from Saccharomyces cerevisiae reveals			
RT	seven open reading frames including the RVS161, ADP1 and PGK genes.";			
RL	Yeast 8:409-417(1992).			
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).			
CC	-1- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; X59720; CAA42328.1; -			
DR	FIR; S19421; S19421.			
DR	GermOnline; 138916; -			
DR	SGD; S0000604; ADP1.			
DR	GO; GO:0005783; C-endoplasmic reticulum; IDA.			
DR	InterPro; IPR003593; AAA_Atpase.			
DR	InterPro; IPR003439; ABC_Transporter.			
DR	Pfam; PF00005; ABC_tran; 1.			
DR	ProDom; PD000006; ABC transporter; 1.			
DR	SMART; SM00382; AAA; 1.			
DR	PROSITE; PS00211; ABC_TRANSPORTER_1; 1.			
DR	PROSITE; PS00893; ABC_TRANSPORTER_2; 1.			
DR	ATP-binding; Transmembrane; Glycoprotein; Transport; Signal.			
KW	SIGNAL 1 25			
FT	CHAIN 26 1049			
FT	NP BIND 423 430			
FT	TRANSMEM 325 345			
FT	TRANSMEM 464 481			
FT	TRANSMEM 794 814			
FT	TRANSMEM 829 849			
FT	TRANSMEM 878 898			
FT	TRANSMEM 910 930			
FT	TRANSMEM 938 958			
FT	TRANSMEM 1001 1021			
FT	TRANSMEM 1025 1045			

```
FT CARBOHYD 50 50 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 221 221 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 815 815 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 935 935 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 960 960 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 971 971 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1049 AA; 117231 MW; ABC9CE54BCDF6A3 CRC64;

Query Match 24.9%; Score 835.5; DB 1; Length 1049;
Best Local Similarity 30.5%; Pred. No. 1.4e-49;
Matches 222; Conservative 134; Mismatches 257; Indels 115; Gaps 18;

QY 1 MSSNVEVPIPVSGQNTNGFPATASNDLKAFTEGAVLSPHNICYVKLKGSLPCRKQVE 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 355 LGSSKSPIRLP-DEDAVNNFLQNEDDTL-----ATLSPENITYSVFSINS-----DQVE 402
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 61 KEILSININGIMKPG-LNAILGPTGGKSSLLDLVLAARKDPSGLSDVLINGAPRP-ANFK 118
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 403 ETVLNEISGIVKQGQILAIMGGGAGKTTLLDILAMKRTKTHVSGSKVINGISMDKSF 462
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 119 CNSGYVQDDVVMGTLTVRENIQFSAALRLATMTNHEKNEIRNVIQELGLDKVADSKV 178
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 463 KIIFGVDQDDFLPTLVTFVTVLNSALLPLKALPEAKKARVYKVLBELRIIDKRII 522
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 179 GTQFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVLLLLKMSKQ-QRT 237
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 523 GNEFDGIGSGGERKRVSIACELVTSPLVFLDEPTSGLDASVANNVIELVLSDDYNT 582
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 238 IIFSIHQPVYSIFKLFDSITLLASRLMHPGPAQELGAFYFESAGYHCEAYNPPAFDLDI 297
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 583 LVLSITHQPSNIFLFDKLVLSGEMVYSGNAKKVSEFLRNEGVTCPDNYNIADYLDI 642
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 298 -----INGDSTAV 305
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 643 TPAGQGGKRRRTENISDLEAGTDTNDIDNTIHQTFTSDGTQREWAHLAHRDEINS 702
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 306 ALNREEDFKATE-----IIFPSKODKPLIEKLABIYVNSFYKETKALHQ-LSGGEKKK 360
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 703 LLRDEEDVEGTGRRGATEIDLNTKLLHDK-----YKDSVYVYAELSQEIIEVLSEGDEN 758
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 361 IT--VPKEISYVTSFCHQLRWKSRFKNLLGNPQASIAQIIVTVVLGVIGAVIYGLKN 418
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 759 VLNGDLPTGQSGAGFTQQLSILNSRFKNRYNPKLLGNLYLTLLLSFLGTLYNVGN 818
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 419 DSTGIQNRAGVLFLLTNQCFSSVAVELFWVEKKLFIHEYISGYRVSSYFLGKLLSDL 478
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 819 DISGFQNRMLGFFFLTYFGVFTFTGLSSPALERIIFIKERSNNYSPLAYYISKIMSEV 878
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 479 LPMTMLPSIIFTCIVYFNLGLKPKADAFVFMFTLMVAVSASSMALATAAQSVSVAT 538
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 879 VPLRVVFPILLSLIVYPMTPGLNKNKDNFAFFKICIGILILNIGLSLEILTGIIFEDLNNSI 938
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 539 LLMTICVFVMFIRPSGLIV---NLTTIASLWSLQYFISIPRYGFTALQHNEF----- 586
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 939 ILSVLVLGSLRGLFINTKNTNVA--FKYLNFSVFYAYESLLINEVTKMLKERK 996
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 587 LGQNF-CPGLNATGNFCNATCTGGEYLVKQGI--DLSPWGLKKNHVALACMIVFLIT 643
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 997 YGLNIEVPG-----ATILSTFGFVVQNLVFDIK-----ILALFNVVFLIM 1036
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 644 AYLKLLFL 651
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1037 GYLALKWI 1044
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

## RESULT 4

```
WHIT DROME
ID WHIT DROME STANDARD; PRT; 687 AA.
AC P10090; Q9V3A2; Q9XY33;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
```

```
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE White protein.
OS W OR EG:BACN3B1.1 OR CG2759.
GN Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Head;
RX MEDLINE=85134865; PubMed=6084717;
RA O'Hare K., Murphy C., Lewis R., Rubin G.M.;
RA "Sequence of a cDNA from the Drosophila melanogaster white gene.";
RA Nucleic Acids Res. 18:1633-1633(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=85134865; PubMed=6084717;
RA O'Hare K., Murphy C., Lewis R., Rubin G.M.;
RA "DNA sequence of the white locus of Drosophila melanogaster.";
RA J. Mol. Biol. 180:437-455(1984).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=21100348; PubMed=11156992;
RA Lukacsovich T., Asztalos Z., Awano W., Baba K., Kondo S., Niwa S.,
RA Yamamoto D.;
RA "Dual-tagging gene trap of novel genes in Drosophila melanogaster.";
RA Genetics 157:727-742(2001).
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN=Berkley; PubMed=107311132;
MEDLINE=20196008; PubMed=107311132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Bens P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasok P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster.";
RA Science 287:2185-2195(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R;
```

RX MEDLINE=20196011; PubMed=10731137;  
 RA Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,  
 RA Barrell B.G., Ferraz C., Vidal S., Brun C., Demailles J., Cadieu E.,  
 RA Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Borkova D.,  
 RA Minana B., Kafatos F.C., Louis C., Siden-Kiamos I., Bolshakov S.,  
 RA Papadogiannakis G., Spanos L., Cox S., Madieno E., de Pablos B.,  
 RA Modolell J., Peter A., Schoettler F., Werner M., Mourikioti F.,  
 RA Beinert N., Dowe G., Schaefer U., Jaeckle H., Bucheton A.,  
 RA Callister D.M., Campbell L.A., Darlamsitou A., Henderson N.S.,  
 RA McMillan P.J., Salles C., Tait E.A., Valenti P., Saunders R.D.C.,  
 RA Glover D.M.;  
 RT "From sequence to chromosome: the tip of the X chromosome of D.  
 RT melangaster";  
 RL Science 287:2220-2222 (2000).  
 RP [6]  
 RP SEQUENCE OF 224-331 FROM N.A.  
 RX MEDLINE=89339145; PubMed=2503416;  
 RA Tearle R.G., Belote J.M., McKeown M., Baker B.S., Howells A.J.;  
 RT "Cloning and characterization of the scarlet gene of Drosophila  
 RT melangaster";  
 RL Genetics 122:595-606 (1989).  
 CC -!- FUNCTION: Part of a membrane-spanning permease system necessary  
 CC for the transport of pigment precursors into pigment cells  
 CC responsible for eye color. White dimerize with brown for the  
 CC transport of guanine and with scarlet for the transport of  
 CC tryptophan.  
 CC -!- SUBUNIT: Heterodimer of white with either brown or scarlet.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC ENBL; X51749; CAA36038.1; --  
 CC ENBL; X02974; CAA26716.1; --  
 CC ENBL; AB028139; BAA78210.1; --  
 CC ENBL; AB003425; AAF45826.1; --  
 CC ENBL; AL333506; CAG55847.1; --  
 CC ENBL; X76202; CAA53795.1; --  
 CC PIR; S08635; FYFFW.  
 CC FlyBase; FBgn003996; w.  
 CC GO; GO:0004888; F:transmembrane receptor activity; NAS.  
 CC GO; GO:0006727; P:chromochrome biosynthesis; IMP.  
 CC InterPro; IPR003593; AAA\_ATPase.  
 CC InterPro; IPR003439; ABC\_transporter.  
 CC InterPro; IPR005284; Pigment\_permease.  
 CC Pfam; PF00005; ABC\_tran; 1.  
 CC ProDom; PD000006; ABC\_transporter; 1.  
 CC SMART; SM00382; AAA; 1.  
 CC TIGRfam; TIGR00955; 3a01204; 1.  
 CC PROSITE; PS0211; ABC\_TRANSPORTER\_1; 1.  
 CC PROSITE; PS0893; ABC\_TRANSPORTER\_2; 1.  
 CC Pigment; ATP-binding; Transmembrane; Transport.  
 KW NP\_BIND 130 137 ATP (BY SIMILARITY).  
 FT TRANSMEM 435 453 POTENTIAL.  
 FT TRANSMEM 465 485 POTENTIAL.  
 FT TRANSMEM 515 533 POTENTIAL.  
 FT TRANSMEM 542 563 POTENTIAL.  
 FT TRANSMEM 576 594 POTENTIAL.  
 FT TRANSMEM 659 678 POTENTIAL.  
 FT CONFLICT 25 29 GDSGA -> LIFEIPYHCRVTAD (IN REF. 2 AND  
 FT CONFLICT 49 49 L -> R (IN REF. 4 AND 5).  
 FT CONFLICT 335 371 VQAQPTNYPADPVOVLAVVPGRETESRDIAKIC ->  
 FT ITLHNSYPAWVSVLPFTIRRTTYRCWPLCPDGRSSPV  
 FT GSPRYG (IN REF. 3).  
 SQ SEQUENCE. 687 AA; 75672 MW; 24AFAD799DE0D396 CRC64;

Query Match 24.28; Score 812; DB 1; Length 687;  
 Best Local Similarity 32.18; Pred. No. 3.3e-46;  
 Matches 210; Conservative 134; Mismatches 251; Indels 60; Gaps 17;  
 QY 5 NVEVFIPVSGQNTGFPATASNDLKAFTEGAVLSFHNICVYRVKLKSGFLPCRPVEKBL 64  
 DB 74 NMDIFGAVNQ-----PGSGWRQLVNRTRGLFCNERHI-----DAPR---KHL 113  
 QY 65 SNINGIKWPG-LNAILQPTGGKSSLLDLAAKDPGSL-----SGDVLINGARPPA-NFK 118  
 DB 114 KNGCVAYPGCELLAVMGSSGAGKTTLLNALAFR-SPOGIQVSPGMRLLNQPVDAKEMQ 172  
 QY 119 CNSGYVVDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVQELGLDKVADSKV 178  
 DB 173 ARCAVVODDLFTGSLTAREHLIFQAVMRPHLTQYRVARDVQVIELSLSKCOHTII 232  
 QY 179 GTQ-FIRGVGGERKRTSIGMELITDPSILFLDEPTTGLDSSSTANAVLLIKMSKQRT 237  
 DB 233 GVPGRVKGSGGERKRLAFASEALTDPPLLCIDEPSTGLDSTFAHSVVQVLKLSQKGT 292  
 QY 238 IIFSIHOPRYSIFKLPDLSLTILASGRLMFHGPAQALGYFESAGYHCEAYNNPADFFLDI 297  
 DB 293 VILTIHQPSSSELFELFKILLMAEGRVAFGLTPEAVDPFSYVGAQCTPNYNPADFYVQV 352  
 QY 298 INGDSTAVLNREDFKATEIIEPSKODKPLIEKLAISIYVNSPYKETKAEHLHOLSGEK 357  
 DB 353 L-----AVVPGRETESR-----DRAIKICDNFAISKVAR-DMEQLLATKN 391  
 QY 358 KKKITVPEKISY--TSFCHLRWVSKRSKKNLGNPQASTAQIIVTWLGLVIGAIYFG 415  
 DB 392 LEKLEQEPENYTKAWFMQFRAVLWRSLVSLKEPLLKVLRIQTMTWAILGLIFLG 451  
 QY 416 LKNDSTGIQNRAGVLFLLTNQCFSSVSA-VELFVVEKKLFIHBYISGYRVSSYFLGKL 474  
 DB 452 QQLTQGVGMINGAIFLFTNMTFQNVFATINVTSELPMVFMREARSLRCDTYFLGKT 511  
 QY 475 LSDLLPMTLPSIIFTCTIVFVLMGLKPKADAFVWMFTLMVAVSASSMALAIAGOSVV 534  
 DB 512 IAE-LPFLVTLVFTALAYPMIGLRAGLVHFFNCLALVTLVANVTSFGVLISCASST 570  
 QY 535 SVATLLMTICFVFMVIFSGLLVNLTTIASLWSLQYFSIPRYGFTALQHNFLQONCPG 594  
 DB 571 SMALSVGPPVLIIPFLFGGFLNSGSPVVLKWSLSWFRYANEGLINQWADVE--PG 628  
 QY 595 -LNATGNPCNVAICTGEEYLVKQIDLSPLGLKWNHVALACMIVILFTAYLKL 648  
 DB 629 EISCTSSN---TTCPSGKVILETLPNSAADPLDYVGLAILIVSRVLAYLAL 679  
 RESULT 5  
 WHIT\_CERCA  
 ID WHIT\_CERCA STANDARD; PRT; 679 AA.  
 AC Q17320;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE White protein.  
 GN W.  
 OS Ceratitis capitata (Mediterranean fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Tephritidae; Tephritidae; Ceratitis.  
 OX NCBI\_TaxID=7213;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9612376; PubMed=8533095;  
 RA Zwiabel L.J., Saccone G., Zacharopoulou A., Besansky N.J.,  
 RA Favia G., Collins F.H., Louis C., Kafatos F.C.;  
 RT "The white gene of Ceratitis capitata: a phenotypic marker for  
 RT germline transformation";  
 RL Science 270:2005-2007 (1995).  
 CC -!- FUNCTION: May be part of a membrane-spanning permease system  
 CC necessary for the transport of pigment precursors into pigment

cells responsible for eye color.  
 -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 -!- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; X89933; CA861398.1; -  
 InterPro; IPR003593; AAA\_ATPase.  
 InterPro; IPR003439; ABC transporter.  
 InterPro; IPR005284; Pigment\_permease.  
 Pfam; PF00005; ABC\_tran; 1.  
 ProDom; PD000006; ABC transporter; 1.  
 SMART; SM00382; AAA; 1.  
 TIGRFAMs; TIGR00955; 3a01204; 1.  
 PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
 PROSITE; PS00893; ABC\_TRANSPORTER\_2; 1.  
 Pigment; ATP-binding; Transmembrane; Transport.  
 NP BIND 121 128 ATP (BY SIMILARITY).  
 FT TRANSMEM 427 445 POTENTIAL.  
 FT TRANSMEM 457 477 POTENTIAL.  
 FT TRANSMEM 507 525 POTENTIAL.  
 FT TRANSMEM 534 555 POTENTIAL.  
 FT TRANSMEM 568 586 POTENTIAL.  
 FT TRANSMEM 651 670 POTENTIAL.  
 FT CARBOHYD 628 628 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 643 643 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 679 AA; 75145 MW; 3F9CB78A835C4C CRC64;

Query Match 23.0%; Score 772; DB 1; Length 679;  
 Best Local Similarity 30.7%; Pred. No. 1.8e-45;  
 Matches 209; Conservative 135; Mismatches 256; Indels 80; Gaps 21;

11 PVSQNTN-GF-----PA-TASN-----DLKAFTEGAV-----LSFHNICRYVK 47  
 30 PYEQSSINGCFCKNYGTLSPSPALTDNLTYSWYNLDVF--GAVHQPSSGSKQLVNRVK 87  
 48 ---LKSGLPCKPVEKEILSNIMKPG-LNALGTGGKSLLDVLAARK----- 97  
 88 GVFCNERHIPAPR---KHLKNDGVAYPGELLAVNGSSGAGKTTLLNASAPRSKGVQI 144  
 98 DPGSLGDLVINGAPRA-PFKCNSGYVQDDVVMGTLTVRENLFQSAALRLATMTNHE 156  
 145 SFSTIR---MLNGHPVDAKEMQARCAVYQDDLFGLSLTAREHLIFQAWVRPHMTQK 201  
 157 KNERINRVQLGLQKVDKSVGTO-FIRGVGGRKSTSGMELITDPSILFLDEPTTG 215  
 202 KVRVDQVQLDLSLQKQNTLIGVPRVGLSGGGERKLAFASEALTDPPLLICDEPTSG 261  
 216 LDSSTANAVLLKMSKGRITIFSIHQPRYSIFKLPDLSLTLLASGLMFHGPQAELG 275  
 262 LDSFWAHSVVQVLLKSLQKGVILTIPHPSSELPFELDKILLMAEGRVALGTFGEAVD 321  
 276 YFESAGYHCEAYNPADPFLDIINGSDTAVALNREED-----FKATEIIFBSKQDKPLIE 330  
 322 FFSYGATCPTNYTPADFYVQVL-----AVVPGREVSERDRVAKICDNFAVGYKVSREMEQ 376  
 331 KLAETVYVNSFFYKETAELHQLSGGKKKITVFKIEISYTSFCHOLRWSKRSPFNLLG 390  
 377 NPKLVKNSGFKGEDE-----NEYTKASWFMQFRAVLWRSLVSK 418  
 391 NPOASTAQIIVTVGLVITGALYFGLKNDSTGIONRAGVLFLTTNQCF-SVSAVELV 449  
 419 EPLLKVRLLQTTMAVLGLIFLQGLTQVGVWNINGAIFLFLTNMTFQNSFATITFT 478  
 450 VEKKLPIFIYSYRVSVSYFLGKLLSLLPLMTPLPSIIFTCIVYFVLGLKPKADRFVM 509  
 479 TELPVMRETRBLRCDFYFLQKIAE-LPLFLVVPFLTAIYPLGLRPGVDHFFTA 537

QY 510 MFTLMVAYSASSMALAIAGQSVVSVATILMTICFVFMIFSGLLVNLTTIASWLSW 569  
 DB 538 LALVTLVANVSTFGYLISCACSSSTMSALSVGPVPIIPFLFGGFFLNSGSPVYFKWLS 597  
 QY 570 YSISRYGFTALQHNHFEIQLGQFCPG-LNATGNPNVATCTGEEVLVKGGIDLSWGLWK 628  
 DB 598 YLSWRYANEGLLINQW--ADVKPGEITCTLSN-----TTCSSSGEVILETINFSASDLFF 651  
 QY 629 NEVALACMIVIFELTIAYLKL 648  
 DB 652 DFIGLALLIVGPRISAYIAL 671

## RESULT 6

WHIT\_LUCCU STANDARD; PRT; 677 AA.  
 ID AC Q05360;  
 DT 01-PEB-1995 (Rel. 31, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE White protein.  
 GN W.  
 OS Lucilia cuprina (Greenbottle fly) (Australian sheep blowfly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Oestroidea;  
 OC Calliphoridae; Lucilia.  
 OX NCBI\_TaxID=7375;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97087158; PubMed=8933176;  
 RA Garcia R.L., Perkins H.D., Howells A.J.;  
 RT "The structure, sequence and developmental pattern of expression of  
 RT the white gene in the blowfly *Lucilia cuprina*.";  
 RL Insect Mol. Biol. 5:251-260(1996).  
 RN [2]  
 RP SEQUENCE OF 490-584 FROM N.A.  
 RX MEDLINE=90264941; PubMed=1971656;  
 RA Elizur A., Vacek A.T., Howells A.J.;  
 RT "Cloning and characterization of the white and topaz eye color genes  
 RT from the sheep blowfly *Lucilia cuprina*.";  
 RL J. Mol. Evol. 30:347-358(1990).  
 CC -!- FUNCTION: May be part of a membrane-spanning permease system  
 CC necessary for the transport of pigment precursors into pigment  
 CC cells responsible for eye color.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; U38899; AAA82057.1; -  
 EMBL; X53265; CAA37365.1; -  
 InterPro; IPR003593; AAA\_ATPase.  
 InterPro; IPR003439; ABC transporter.  
 InterPro; IPR005284; Pigment\_permease.  
 Pfam; PF00005; ABC\_tran; 1.  
 ProDom; PD000006; ABC transporter; 1.  
 SMART; SM00382; AAA; 1.  
 TIGRFAMs; TIGR00955; 3a01204; 1.  
 PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
 PROSITE; PS00893; ABC\_TRANSPORTER\_2; 1.  
 KW Pigment; ATP-binding; Transmembrane; Transport.  
 NP BIND 119 126 ATP (POTENTIAL).  
 FT TRANSMEM 431 451 POTENTIAL.  
 FT TRANSMEM 456 476 POTENTIAL.  
 FT TRANSMEM 506 526 POTENTIAL.  
 FT TRANSMEM 534 554 POTENTIAL.



FT TRANSEM 563 583 POTENTIAL.  
 FT TRANSEM 647 667 POTENTIAL.  
 SQ SEQUENCE 677 AA; 75365 MW; D16FC11C97EED51D CRC64;  
 Query Match 22.5%; Score 754.5; DB 1; Length 677;  
 Best Local Similarity 31.4%; Pred. No. 2.9e-44;  
 Matches 194; Conservative 123; Mismatches 226; Indels 75; Gaps 18;

QY 5 NVVFTVPSQNTNGFPATASNDLKAFTEGAVLSFNICVRVK---LKSGLPCRPVEK 61  
 DB 63 NLDFVEVHPGSEN-----WKQVNRVGVFCNERHHP--KP-RK 99  
 QY 62 EILSNINGIMKPG-LNAILGPTGGKSSLDVLAARK-----DPSGLSGDVLINGAPP 114  
 DB 100 HLIKCGVAYPGBELLAVMGSSGAGTKLLNALAFRSARGVQISPSVR--MLNGHPVD 156  
 QY 115 A-NFKNSGYVQDDVVMGTLTVRENLOFSAALRLATTNHEKNERINRVQELGDKV 173  
 DB 157 AKEMQARCAVYQDDDLFIGSLTAREHLIQAATVRMPRTMTQKQQRVDQVQDLSLIK 216  
 QY 174 ADSKVGTO-FIRGVSGERKRTSIGMELITDPSILFDEPTTGLDSTANAVLLIKRMS 232  
 DB 217 QNTIIGVGRVKGSGERKRLAFASEALTDPLLCIDPTSGLDSEFMAASVVQVLKLS 276  
 QY 233 KQRTIIFSIHQPRYSIFKLPDLSLTLASGRLMFHGPAQALGYFESAGHCEAVNPNAD 292  
 DB 277 QRGKTIVLTTHQPSSELFELDKILLMAEGRVAFGLTPVEAVDPFFIGAQCPNTNPNAD 336  
 QY 293 PFDIINGDSTAVALNREEDFKATEIIEPSKQDKPIEKLAETVYNSVFVKETKAELHOL 352  
 DB 337 FYVQVL-----AVVPGRE-----IESDRISKIDCNFAVGKVSREMEQNFQKI 379  
 QY 353 S-----GEEKKKKTIVFKEISYTTSPCHLRWWSKRSFKMLLGNPQASIAQIIVTVVLGV 408  
 DB 380 AAKTDGLQKDDETILYKASWFTQF-----RAIMWRSWISTLKEPLLVKVLQITTMVAVL 435  
 QY 409 IGALYFGLKNDSTGIONRAGVLFELTNTQCFSSVSAV-ELFVVEKLFTHIEVSGYRVS 467  
 DB 436 IGLIFLNPQMTQGVWMINGAIFLPLTNMTFQNVFAVINFTSELVFRTRSRLYRCD 495  
 QY 468 SYFLGKLLSLLPMTLPSIIFTIVFVLMGLKPKADAFVMMFTMLMVAYSASSMALAI 527  
 DB 496 TYFLGKTLAE-LPLFLVPELFAIAYPMIGLRPGITHFLSALALVTLVANVSTSPGYLI 554  
 QY 528 AAGOSVSVATLL---MTICFVFMIFSLGLLVNLTIASLWSLQVFSFRICFTALQHN 584  
 DB 555 SCASTSTSMALSGVPLTIPTF---LLFGVFLNSGVVYFVKWLSYFSFRYANEGLLIN 611  
 QY 585 EFL---GQNFPCPLNAT 598  
 DB 612 QWADVQPGEITCTSTNTT 629

RESULT 7  
 WHIT ANOGA  
 ID WHIT ANOGA STANDARD; PRT; 695 AA.  
 AC Q27256; Q17006;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE White protein.  
 GN W.  
 OS Anopheles gambiae (African malaria mosquito).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.  
 OC NCBI\_TaxID=7165;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Suakoko / G3;  
 RX MEDLINE=96423158; PubMed=8825759;  
 RA Besansky N.J., Bedell J.A., Benedict M.Q., Mukabayire O., Hilfiker D.,  
 RA Collins F.H.;  
 RT "Cloning and characterization of the white gene from Anopheles

gambiae.";  
 Insect Mol. Biol. 4:217-231 (1995).  
 !- FUNCTION: May be part of a membrane-spanning permease system  
 necessary for the transport of pigment precursors into pigment  
 cells responsible for eye color.  
 !- SUBCELLULAR LOCATION: Integral membrane protein.  
 !- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.ebi.ac.uk/announcements/>  
 or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

EMBL; U29486; AAC46995.1; -;  
 EMBL; U29485; AAC46994.1; -;  
 EMBL; U29484; AAC47423.1; -;  
 InterPro; IPR003593; AAA\_ATPase.  
 InterPro; IPR003439; ABC\_transporter.  
 InterPro; IPR008965; Cellul\_bind.  
 InterPro; IPR005284; Pigment\_permease.  
 Pfam; PF00005; ABC\_tran; 1.  
 ProDom; PD000006; ABC\_transporter; 1.  
 SMART; SM00382; AAA; 1.  
 TIGRFAMs; TIGR00955; 3a01204; 1.  
 PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
 PROSITE; PS00893; ABC\_TRANSPORTER\_2; 1.  
 KW Pigment; ATP-binding; Transmembrane; Transport.  
 FT NP\_BIND 133 140 ATP (POTENTIAL).  
 FT NP\_BIND 288 295 ATP (POTENTIAL).  
 FT TRANSMEM 444 464 POTENTIAL.  
 FT TRANSMEM 474 494 POTENTIAL.  
 FT TRANSMEM 524 544 POTENTIAL.  
 FT TRANSMEM 552 572 POTENTIAL.  
 FT TRANSMEM 581 601 POTENTIAL.  
 FT TRANSMEM 659 689 POTENTIAL.  
 FT CARBOHYD 472 472 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 100 100 N -> S (IN REF. 1; AAC47423).  
 FT CONFLICT 691 693 SRS -> YAR (IN REF. 1; AAC47423).  
 SQ SEQUENCE 695 AA; 77218 MW; ER8B9517239E2961 CRC64;

Query Match 22.4%; Score 751; DB 1; Length 695;  
 Best Local Similarity 30.0%; Pred. No. 5.3e-44;  
 Matches 206; Conservative 129; Mismatches 273; Indels 78; Gaps 20;

QY 6 VEVFIPVSQNTNG-FP-----ATASNDLKAFTEGAVLSFNICVRVKLK 49  
 DB 50 IQVWPKSYSGVKGQIPQICERLTVYWEIDVFGAPTQKRP-REFLCRLNCCTR--QR 106  
 QY 50 SGFLPCRPVEKEILSNINGIMKPG-LNAILGPTGGKSSLDVLAARKOPS-GLSGDVL 107  
 DB 107 KDFNP-----RHLTKNVTGVAKSGELLAVMGSSGAGTKLLNALAFRSPGVKISPNV 161  
 QY 108 --INGAPRPA-NFKNSGYVQDDVVMGTLTVRENLOFSAALRLATTNHEKNERINRV 164  
 DB 162 RALNGVPVNAEQRLARCAVYQDDDLFIPLSLTTRHELLFOAMLRMGDRDVPASVKCHRVQEV 221  
 QY 165 IQELGLDKVADSKVGTQ-FIRGVSGERKRTSIGMELITDPSILFDEPTTGLDSTANA 223  
 DB 222 IQELSLVKCADTIIGAPGRKGLSGERKRLAFASETLTDPHLLCCDEFTSGLDSEFMAHS 281  
 QY 224 VLLILKRMKSGKRTIIFSIHQPRYSIFKLPDLSLTLASGRLMFHGPAQALGYFESAGYH 283  
 DB 282 VLQVLKGMWAKGKTIILTIHQPSSELYCLFDKILLVAEGRVAFGLSPYQSAEFFSQLGIP 341  
 QY 284 CEAVNPNADPFLDINGDSTAVALNREEDFKATEIIEPSKQD--KPLIEKLAETVYNSVF 341  
 DB 342 CPPNYNADFFVQML-----AIAPAKEAEORCDMKIKICDSFAVSPI 382  
 QY 342 YKETKAELHQLSGGEEK-----KKITVPKEISYTTSPCHLRWWSKRSFKMLLGNPQASIA 397

Db 383 AREVLETASVAGKMGDEPYMLQOEVGSGTGRSSWMTQFCILWRGSLVLDKPMVLKV 442  
 QY 398 QIIVTVGLVIGAIYFGLKNDSTGIONRAGVLFELTNQCFSSVAV-ELFVVEKKLFI 456  
 Db 443 RLQQTAMVATLIGISYFQOVLQDQGVNINGSLEFLFTNMTFQNVFVAVINVSABLPVL 502  
 QY 457 HEIISGYRVSSYFGLKLSLLPMTLPSIIFTCIVYFVGLGPKADAFVMMFTLMV 516  
 Db 503 REKRSRLYRVDYFLGKTIAE-LPLFIAPVFTSITVPMIGLRTGATHYTLTLFIVTLV 561  
 QY 517 AYSASSMALAIAGOSVSVATLLMTICFVFMFISGLLVNLTIAIWSLWLOFSPRY 576  
 Db 562 ANVSTFGLVISCASSISMAISVGPVVIIFLFGFPLNSASVPAYFKYLSVLSWRY 621  
 QY 577 OPTALQHNEL----GQNFPCPLNATGNPCNYATCTGEEYLKQ-GTIDSPMGLKWHV 631  
 Db 622 ANEALLINQWSVVDGEIACTRAN------VTCPESEIILETFNFRVEDPAL- 667  
 QY 632 ALACM--IVIFLTIAYLKLFKKYS 655  
 Db 668 DIACLFALIVLFRGLALCLMLRSRS 693

RESULT 8  
 ID WHIT\_ANOAL STANDARD; PRT; 709 AA.  
 AC Q16928;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE White protein.  
 GN W.  
 OS Anopheles albimanus (New world malaria mosquito).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Neoptera; Culicoidae; Anopheles.  
 OX NCBI\_TaxID=7167;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN:TeCo.  
 RX MEDLINE=98726549; PubMed=9465401;  
 RA Ke Z., Benedict M.Q., Cornet A.J., Besansky N.J., Collins F.H.;  
 RT "The Anopheles albimanus white gene: molecular characterization of  
 the gene and a spontaneous white gene mutation.";  
 RL Genetics 101:87-96(1997).  
 CC -!- FUNCTION: May be part of a membrane-spanning permease system  
 CC necessary for the transport of pigment precursors into pigment  
 CC cells responsible for eye color.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; L76302; AAA88240.1; -  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR InterPro; IPR005284; Pigment\_permease.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR TIGRFAMs; TIGR00955; 3a01204; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; FALSE\_NEG.  
 DR PROSITE; PS00893; ABC\_TRANSPORTER\_2; 1.  
 KW Pigment; ATP-binding; Transmembrane; Transport.  
 FT NP\_BIND 136 143 ATP (POTENTIAL).  
 FT NP\_BIND 292 299 ATP (POTENTIAL).  
 FT TRANSMEM 457 475 POTENTIAL.

FT TRANSMEM 487 507 POTENTIAL.  
 FT TRANSMEM 537 555 POTENTIAL.  
 FT TRANSMEM 564 585 POTENTIAL.  
 FT TRANSMEM 598 616 POTENTIAL.  
 FT TRANSMEM 681 700 POTENTIAL.  
 FT CARBOHYD 485 485 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 658 658 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 709 AA; 79052 MW; FBD76D5C69D3BECA CRC64;  
 Query Match 22.0%; Score 736.5; DB 1; Length 709;  
 Best Local Similarity 30.1%; Pred. No. 5.4e-43;  
 Matches 210; Conservative 135; Mismatches 265; Indels 87; Gaps 24;  
 QY 6 VEVFPVSGQNTNG-FPAT------ASNDLKAPTEGAV--LSFHNLCYRV-----KLSK 50  
 Db 51 IQWRPKSYSGYKGIQAQDLTYTWREIDVFGAAIDKGRPLCSLRHCFTRQRLVK 110  
 QY 51 GLPLCKPKVEKILSNINGIMKPG-LNAILGPTGGKSLDLDLAARKDPS-GLSGDVL- 107  
 Db 111 DFNP-----RKHLKNVTGVARSSELLAVMGSSGAGKTTLLNELAFRPPGVKISPAIR 165  
 QY 108 -INGAPRPA-NPKCMSGVVQDDVVMGTLTVRENIQPSAALRL-----ATTNNEKNER 160  
 Db 166 TLNGVDTVAEOMRARCAYVQDDLFIPSLTTKEHLMFQAMLEMRDVFATPIKMH---R 221  
 QY 161 INRVQELGLDVADSKVGTQ-FIRGVGGERKRTSIGMELITDPSIILFDPTTGLDSS 219  
 Db 222 VDEVLQELSLVKADTIIVAGRVKLSGGERKRTAFRSEITLDPHLLCDEPTSLDSF 281  
 QY 220 TANAVALLKRMKSGRTIIFSIHQPRYSIFKLFDLSLTLLASGRMLFHGPAQALGYFES 279  
 Db 282 MAQSVLQVLKGMAMKGTIILTIHQPSSELYCLFDRILLVARG-VAFGLSPYQSADFFSQ 340  
 QY 280 AGVHCAYNPNADPFLDIINGSTVALNRDEDKATEIIEPSKODKPLIEKLAELVYNS 339  
 Db 341 LGIPCPNPVNPADFYVQML-----ALAPNKETERET-----IKKICSFAYS 383  
 QY 340 SFYK---ETKAEHLQLSGGE-KKKKITV-----FKEISYTTSFCHQLRWVSKRSF 385  
 Db 384 PIARDIETASQVNGDGGIELTRKHTTDPYFLQPMEGVDSTGYRASWMTQFCILWRSW 443  
 QY 386 XNLLGNPQASIAQIIVTVGLVIGAIYFGLKNDSTGIONRAGVLFELTNQCFSSVAV 445  
 Db 444 LSVLKDPMLVKVRLLOQTAVASLIGSYFQVLDQGVNINGSLEFLFTNMTFQNVFVAV 503  
 QY 446 -ELFVVEKKLFIHEIISGYRVSSYFGLKLSLLPMTLPSIIFTCIVYFVGLGPKAD 504  
 Db 504 INVFSALPVLREKESRLYRVDYFLGKTIAE-LPLFIAPVFTSITVPMIGLKAATS 562  
 QY 505 APTVMFTLMVAVSASSMALAIAGOSVSVATLLMTICFVFMFISGLLVNLTIAIWS 564  
 Db 563 HYLTLTIFIVTLVANVSTFGLYDISCASSISMAISVGPVVIIFLFGFPLNSASVPAY 622  
 QY 565 LSWLOFSPRYGFTALQHNEL----GQNFPCPLNATGNPCNYATCTGEEYLKQKQID 620  
 Db 623 FKLSVLSWRYANEALLINQWADHEDGEIGCTRANVT-----CPASGEIILETFNER 675  
 QY 621 LSPWGLKWHVALACM--IVIFLTIAYLKLFKKYS 655  
 Db 676 VEDFAL-----DIGCLFALIVLFRGLALCLMLRSRS 707

RESULT 9  
 ABGI\_MOUSE STANDARD; PRT; 666 AA.  
 ID ABGI\_MOUSE  
 AC Q64343;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE ATP-binding cassette, sub-family G, member 1 (White protein homolog)  
 DE (ATP-binding cassette transporter 8).  
 GN ABCG1 OR ABCG8 OR WH1L.  
 OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10096;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97186700; PubMed=9034316;  
 RA Croop J.M., Tiller G.B., Fletcher J.A., Lux M.L., Raab E.,  
 RA Goldenson D., Son D., Arciniegas S., Wu R.,  
 RT "Isolation and characterization of a mammalian homolog of the  
 RT Drosophila white gene";  
 RL Gene 185:77-85(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DBA/2;  
 RA MEDLINE=96359154; PubMed=87031120;  
 RX Savary S., Denizot F., Luciani M.-F., Mattei M.-G., Chimini G.,  
 RT "Molecular cloning of a mammalian ABC transporter homologous to  
 RT Drosophila white gene";  
 RL Mamm. Genome 7:673-676(1996).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21092576; PubMed=11162488;  
 RA Lorkowski S., Rust S., Engel T., Jung E., Tegeltkamp K., Galinski E.A.,  
 RA Asmann G., Cullen P.,  
 RT "Genomic sequence and structure of the human ABCG1 (ABC8) gene";  
 RL Biochem. Biophys. Res. Commun. 280:121-131(2001).  
 RN [4]  
 RP INDUCTION, AND PROBABLE FUNCTION.  
 RX MEDLINE=20261604; PubMed=10799558;  
 RA Venkateswaran A., Repa J.J., Lobaccaro J.-M.A., Bronson A.,  
 RA Mangelsdorf D.J., Edwards P.A.,  
 RT "Human white/murine ABC8 mRNA levels are highly induced in  
 RT lipid-loaded macrophages. A transcriptional role for specific  
 RT oxysterols";  
 RL J. Biol. Chem. 275:14700-14707(2000).  
 RN [5]  
 RP REVIEW.  
 RX MEDLINE=21474438; PubMed=11590207;  
 RA Schmitz G., Langmann T., Heimerl S.,  
 RT "Role of ABCG1 and other ABCG family members in lipid metabolism.";  
 RL J. Lipid Res. 42:1513-1520(2001).  
 CC -I- FUNCTION: Transporter involved in macrophage lipid homeostasis. Is  
 CC an active component of the macrophage lipid export complex. Could  
 CC also be involved in intracellular lipid transport processes. The  
 CC role in cellular lipid homeostasis may not be limited to  
 CC macrophages.  
 CC -I- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
 CC -I- TISSUE SPECIFICITY: Expressed mainly in brain, thymus, lung,  
 CC adrenals, spleen and placenta. Little or no expression in liver,  
 CC kidney, heart, muscle or testes.  
 CC -I- INDUCTION: Strongly induced in macrophage cell line RAW264.7  
 CC during cholesterol influx. Induction is mediated by the liver X  
 CC receptor/retinoid X receptor (LXR/RXR) pathway.  
 CC -I- SIMILARITY: Belongs to the ABC transporter family. ABCG (white)  
 CC subfamily.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U34920; AAB47738.1; --  
 DR EMBL; Z49745; CAA88636.1; --  
 DR EMBL; AF323659; AAK27442.1; --  
 DR MGD; MGI:107704; Abcg1.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR InterPro; IPR005284; Pigment\_permease.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.

DR SMART: SMO0382; AAA; 1.  
 DR TIGSPAMS; TIGR00955; Jao1204; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
 DR PROSITE; PS00893; ABC\_TRANSPORTER\_2; 1.  
 KW Transport; Lipid transport; ATP-binding; Transmembrane.  
 FT DOMAIN 1 414 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 415 433 POTENTIAL.  
 FT DOMAIN 434 444 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 445 465 POTENTIAL.  
 FT DOMAIN 466 494 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 495 513 POTENTIAL.  
 FT DOMAIN 514 521 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 522 543 POTENTIAL.  
 FT DOMAIN 544 555 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 556 574 POTENTIAL.  
 FT DOMAIN 575 637 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 638 657 POTENTIAL.  
 FT DOMAIN 658 666 CYTOPLASMIC (POTENTIAL).  
 FT NP\_BIND 118 125 ATP (POTENTIAL).  
 SQ SEQUENCE 666 AA; 74033 MW; EDDGAFBD43950B6 CRC64;  
 Query Match 21.3%; Score 714.5; DB 1; Length 666;  
 Best Local Similarity 28.9%; Pred. No. 1.6e-41;  
 Matches 198; Conservative 145; Mismatches 244; Indels 97; Gaps 24;  
 QY 3 SSVVEVPIVSOQNTGFPATASNDLKAFTE-----GAV-LSFHNICRYK---- 47  
 DB 38 SSNVD---EVEDLLNGHLKKVDNN---FTEAQRFSLPRAAVNTEFKDLSYVPEGPW 91  
 QY 48 -LKSGLPCRPKEIKLSNINGIMKPG-LNALIGPTGGGKSLDLVLAARKDPSGLSGD 105  
 DB 92 WKXGKY-----KTLKLGISGKNSGELVINGPSGACKSLMNLVAGYRE-TGMKGA 142  
 QY 106 VLINGAPRPAN-PKCSGYVQDDVVGVTITVRENLOFSALRLATTTWTHEKNERINRV 164  
 DB 143 VLINGMPRLRCERKVCYIMQDDMLPLHLTVQEMMVSAHLKQE--KDEGRREWKYEI 200  
 QY 165 IQELGLDKVADSKVGTQFIRGVSGGERKRTSIGNELITDPSILFDEPTTGLDSSTANAV 224  
 DB 201 LTALGLPLCANTRTGS-----LSGQQRKRLAIALELVNPPVWFDEPTGLDSASCFQV 255  
 QY 225 LLLAKRMKQKQRTIIFSHQPRYSIFKLFDSLTLASGRMLFHGPAQALGYFESAGVHC 284  
 DB 256 VSLMKGLAQGQSIVCTIHQPSAKLFFDLQYVLSQGCYVRGKVSNLVPLFRLDLGLNC 315  
 QY 285 EAYNNDPFLDLINGP-----STAVALNRE-----EDFKATEIIEPSKQDKPLIEK 331  
 DB 316 PTYNPADFVNEVASGEYGDQNSRLVRAVREGMCDADYKDLGGDTDVNFLEWHPAEED 375  
 QY 332 LAEIVVNSSFYKTKAEHLQLSGGKCKKITVFKESITTSFCHQLRWKSKRSPKLLGN 391  
 DB 376 SASMEGCHSF-----SASCLTQFC-----ILPKRTFLSIMRD 407  
 QY 392 PQASIAQIIVTVGLVIGAIYFGLKNDSTGIQNRAGVLPFLTNQCFSSVSVEL-FVV 450  
 DB 408 SVLTHLRITSHIGGLIGLIGLIGLIGNEAKKVLNSGFLFSLFMLFPAALMPTVLPPL 467  
 QY 451 EKCLFIHEYISGYRVSSYFLGKLLSDLLPMTLMPSIIFTCIVYFMGLGPKADAFVPM 510  
 DB 468 EMSVFLREHLNYSLKAYLAKTMAD-VPPQIMFPVAYGSIYVWMTS-QPSDAVRFLVF 525  
 QY 511 FTL-MWVAYSASSMALAIACQSVSVATLMTCTCFVMMIFPSGLLNLITIASLWSLQ 569  
 DB 526 AALGTMTSLSVAQSLGLIGASTSLQVATFVGPTAIPVLLFSGFFVFDTPRAYLQMS 585  
 QY 570 YFSIPRYGFTALQHNELFQNGFCPLNATGNPCNYA-TC--TGEELVKGQIDLSPWGL 626  
 DB 586 YISVRYGFEGVLSIY-----GLDREDLH-CDIAETCHFKQSEAILRE-LDVENAKL 636  
 QY 627 WKNHVALACMTVIFETAYLKLFF 650  
 DB 637 YLDFIVLGIFFISLRIAYFLVRY 660

DR	InterPro; IPRO03593; AAA_Atpase.
DR	InterPro; IPRO03439; ABC_transporter.
DR	Pfam; PF00005; ABC_tran; 1.
DR	ProDom; PD00006; ABC_transporter; 1.
DR	SMART; SMO0382; AAA; 1.
DR	PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR	PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
KW	ATP-binding; Glycoprotein; Transmembrane; Transport.
FT	DOMAIN 1 393 CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM 394 414 1 (POTENTIAL).
FT	DOMAIN 415 425 EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM 426 446 2 (POTENTIAL).
FT	DOMAIN 447 472 CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM 473 493 3 (POTENTIAL).
FT	DOMAIN 494 503 EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM 504 524 4 (POTENTIAL).
FT	DOMAIN 525 532 CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM 533 553 5 (POTENTIAL).
FT	DOMAIN 554 617 EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM 618 638 6 (POTENTIAL).
FT	DOMAIN 639 646 CYTOPLASMIC (POTENTIAL).
FT	NP BIND 102 109 ATP (POTENTIAL).
FT	CARBOHYD 422 422 N-LINKED (GLCNAC . .) (POTENTIAL).
SQ	SEQUENCE 646 AA; 71895 MW; 9CCB6E150772611 CRC64;
 Query Match 21.1%; Score 708; DB 1; Length 646; Best Local Similarity 27.7%; Pred. No. 4.4e-41; Matches 176; Conservative 141; Mismatches 247; Indels 72; Gaps 16;	
QY	37 ISFHNICVRVKLGSGFLPC-RKPEVEILLNSNG-IMKPGLNAIIGPTGGGKSLLDLVLA 94
DB	61 IEFVELSV--REG--PCWRKGKYTLIKLSGRFCRELINGSPSGAGSKTFPMILA 116
QY	95 AKDPSSLGSVDVLINGAPRP-AFPKNCSGYVVODDVVMGTGLTVRENLFOSALRLATTMT 153
DB	117 GYRE-SGMKGQILVNRPRELRTFRMSVCYIQDDMLPHLTLEAMMVVSANLKISE--K 173
QY	154 NHEKNERINRVIQELGDKVADSKVGTPQIRGVSGGERKRTSIGMELITDSILFLDEPT 213
DB	174 QEVKKELTEILTALGLMSCSHRTAL-----LSGQRKLAIAEULVNVPWFVFDEPT 228
QY	214 TGLSDSTANAYLLLKRMSKQGRTHIFSIHQPRYSIFKLFDSLTLLASGRLMFHGAQEA 273
DB	229 SGLDSASCQVQVSLMKSLAQGRTHICTIHQPASKLFEMFDKLYLSQGQCIFKGVNTL 288
QY	274 LCYPESAGYCEAYNPADEFLOIINGD-----STAVALNREEDFKATEII 319
DB	289 IPLYKGLGLHCPETHNPADPIIIVASGEYGDNLNPMLEFRAVQNGLCAAMEAKSSSEPKNEVP 348
QY	320 EFSKQDKPLIEKLAEIIVYNSSFYKETAELHQLSOGKEKKKITVPEKISYTTTSFCHQLRW 379
DB	349 APCPPCPEVDPI-----SSH-----TFATSTLTQPCI 376
QY	380 VKRSFKNNLGNPOASTAQIIVTVLGLVIGAIYFGLNKNDSTGIQNRAGVLFFLTNQCF 439
DB	377 LFKRTFSLISDTVLTLELRMSHWVTGLVGLYLTHIGDDASKVFNNTGCCFFSMLFMF 436
QY	440 SSVSAVEL-FVVEKKLFIHYISGYRYRVSSYFLGKLLSDLPMTPMLPSIITCTIVYFMLG 498
DB	437 AALMPTVLTLPLEVAFVMEHLNVWYSLRAYILAKTWAD-VPFQVVCVPWYCIVYVMWG 495
QY	499 LKPKADAFVNMFTLMVAVSASMALAIAAGOSVVSVAITLMTTCFVFWMIIFSGLLVNL 558
DB	496 QPATSRTELLFSALATATALVAQSGLLIGAANSLSQVATFVGPTAITPLLFSGFFVFSF 555
QY	559 TTIASWLSWLQYFSIPRGYATLQHNEFLQNCFCPLGNATGNPNQYATCT-GSEYLVKQ 617
DB	556 KTIPIYQLWSYLSVAYEGEVILITY-----GWE-RGDLTCLBERCQPREPQSILR 607
QY	618 GIDISPGLWKQNHVALACMIVFEITAYLKLFLKK 653
DB	608 ALDVEDAKLYMDFLVGIFFLALRELLAYLVRYRK 643

```

RESULT 11
ID ABG1_HUMAN STANDARD; PRT; 678 AA.
AC P45844; Q9BXK6; Q9BXK7; Q9BXK8; Q9BXK9; Q9BXL0; Q9BXL1; Q9BXL2;
AC Q9BXL3; Q9BXL4;
DT 01-NOV-1995 (Rel. 32, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE ATP-binding cassette, sub-family G, member 1 (White protein homolog)
DE (ATP-binding cassette transporter 8).
DE ABCG1 OR ABC8 OR WH1.
GN ABCG1 OR ABC8 OR WH1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=20289799; PubMed=10830953;
RC TISSUE=Retina;
RX Chen H.M., Rossier C., Laliet M.D., Lynn A., Chakravarti A.,
RA Perrin G., Antonarakis S.E.;
RT "Cloning of the cDNA for a human homologue of the Drosophila white
RL gene and mapping to chromosome 21q22.3.";
RL Am. J. Hum. Genet. 59:66-75(1996).
RN [2]
RP MEDLINE=20289799; PubMed=10830953;
RX Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y.,
RA Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,
RA Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordieck G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Rieseemann L., Dgand E.,
RA Wehmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehrach H., Reinhardt R., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 21.";
RL Nature 405:311-319(2000).
RN [3]
RP MEDLINE=20408883; PubMed=10950923;
RX Berry A., Scott H.S., Kudoh J., Talior I., Korostishevsky M.,
RA Wattenhofer M., Guipponi M., Barras C., Rossier C., Shibuya K.,
RA Wang J., Kawasaki K., Asakawa S., Minoshima S., Shimizu N.,
RA Antonarakis S.E., Bonne-Tamir B.;
RT "Refined localization of autosomal recessive nonsyndromic deafness
RL structure, and exclusion of six known genes in the region.";
RL Genomics 68:122-29(2000).
RN [4]
RP MEDLINE=21192304; PubMed=11279031;
RX Porsch-Oezcuermmez M., Langmann T., Heimerl S., Borsukova H.,
RA Kaminski W.E., Drobnik W., Honer C., Schumacher C., Schmitz G.;
RT "The zinc finger protein 202 (ZNF202) is a transcriptional repressor
of ATP binding cassette transporter A1 (ABCA1) and ABCG1 gene
expression and a modulator of cellular lipid efflux.";
RL J. Biol. Chem. 276:12427-12433(2001).
RN [5]
RP MEDLINE=21092576; PubMed=11162488;
RX Lorkowski S., Rust S., Engel T., Jung E., Tegelkamp K., Galinski E.A.,
RA Asmann G., Cullen P.;
RT "Genomic sequence and structure of the human ABCG1 (ABC8) gene.";
RL Biochem. Biophys. Res. Commun. 280:121-131(2001).
RN [6]
RP MEDLINE=97196700; PubMed=9034316;
RX Tissue=Petal brain;
MEDLINE=97196700; PubMed=9034316;
RA

```

```

RA Croop J.M., Tiller G.E., Fletcher J.A., Lux M.L., Raab E.,
RA Goldenson D., Arciniegas S., Son D., Wu R.;
RT "Isolation and characterization of a mammalian homolog of the
RL Drosophila white gene.";
RL Gene 185:77-85(1997).
RN [7]
RP MEDLINE=20261604; PubMed=10799556;
RX Venkateswarar A., Repa J.J., Lobaccaro J.-M.A., Bronson A.,
RA Mangelsdorf D.J., Edwards P.A.;
RT "Human white/murine ABC8 mRNA levels are highly induced in
RL lipid-loaded macrophages. A transcriptional role for specific
RL oxysterols.";
RL J. Biol. Chem. 275:14700-14707(2000).
RN [8]
RP MEDLINE=20105556; PubMed=10639163;
RX Klucken J., Buechler C., Orso E., Kaminski W.E.,
RA Porsch-Oezcuermmez M., Liebisch G., Kapinsky M., Diederich W.,
RA Drobnik W., Dean M., Allikmets R., Schmitz G.;
RT "ABCG1 (ABC8), the human homolog of the Drosophila white gene, is a
RL regulator of macrophage cholesterol and phospholipid transport.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:817-822(2000).
RN [9]
RP MEDLINE=21474438; PubMed=11590207;
RX Schmitz G., Langmann T., Heimerl S.;
RT "Role of ABCG1 and other ABCG family members in lipid metabolism.";
RL J. Lipid Res. 42:1513-1520(2001).
CC -!- FUNCTION: Transporter involved in macrophage lipid homeostasis. Is
CC an active component of the macrophage lipid export complex. Could
CC also be involved in intracellular lipid transport processes. The
CC role in cellular lipid homeostasis may not be limited to
CC macrophages.
CC -!- SUBUNIT: May form heterodimers with several heterologous partners
CC of the ABCG subfamily.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Predominantly
CC localized in the intracellular compartments mainly associated with
CC the endoplasmic reticulum (ER) and Golgi membranes.
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=7;
CC Comment=Additional isoforms seem to exist;
CC Name=1;
CC IsoId=P45844-1; Sequence=Displayed;
CC Name=2; Synonyms=J;
CC IsoId=P45844-2; Sequence=VSP_000047, VSP_000051;
CC Name=3; Synonyms=ABDE;
CC IsoId=P45844-3; Sequence=VSP_000048, VSP_000051;
CC Name=4; Synonyms=G;
CC IsoId=P45844-4; Sequence=VSP_000051;
CC Name=5; Synonyms=P;
CC IsoId=P45844-5; Sequence=VSP_000049, VSP_000051;
CC Name=6; Synonyms=Hi;
CC IsoId=P45844-6; Sequence=VSP_000046, VSP_000051;
CC Name=7; Synonyms=C;
CC IsoId=P45844-7; Sequence=VSP_000050, VSP_000051;
CC -!- TISSUE SPECIFICITY: EXPRESSED IN SEVERAL TISSUES.
CC -!- INDUCTION: Strongly induced in monocyte-derived macrophages during
CC cholesterol influx. Conversely, mRNA and protein expression are
CC suppressed by lipid efflux. Induction is mediated by the liver x
CC receptor/retinoid x receptor (LXR/RXR) pathway.
CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)
CC subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; X91249; CAA62631.1; ALT_INIT.
CC

```

173	QY	VADSKVGQTFRTGVSGBERKRTSTGMBLIITDPSILFDDEPTCLDSDSTANAVYLLIKRMS	232
209	Db	CANTRTGS-----LSGQKRLAIALSLVNNPVMFFDEPTSLDSSACFQVVSMLKXGLA	263
233	QY	KOGRTIIFSIHQPRYSIFPKLFDLSLTLIASRLMFHGPQAQEALGYFSSAGHYCEAYNNPAD	292
264	Db	QGRSIICTIHQPSAKLFEFLDQLVYLSQGVYRKVCNLYPFLRDLGLNCTPYENPAD	323
293	QY	PFDLIINGDSTAVALNREEDFKATEIIEPSQDKPLIEKLAIEIYVNSVFKET--KAEI-	349
324	Db	FYMEVASG-----EYGDQNSRLVRAVREGMCDSDHKRDLGGDAEYN	366
350	QY	-----HOLSGGEKK-KKIITVFKXEISYTSFCH-----QLRWVSXKRSFKNLIGNPQASI	396
365	Db	PFLMHPSPSEYVKOTKELKGLKEDSSMEGCHSFASCLTQFCILFKRTFLSIMRDSVLTH	424
397	QY	AQIIVTVVLGLVIGNIYFGLKNDSTGTQNRAGVLFELTTNQCSVSVSAVEL-FVWEKFLF	455
425	Db	LRITSHIGIGLIGLYLIGINEAKVLSNGSGLFFSMFLMFAALMPTVTLPTFLEMGVF	484
456	QY	IHEYISGVYRVSSYFELGKLLDGLPMTMLPSIITFCIYFEMGLGPKADAFFVMMETL-M	514
485	Db	LEHLYNYSLKAYLYATMAD-VPFQIMPPVACSIYVWMTS-QPSDAVRFVLFAALGT	542
515	QY	MVAYSASSNALAIAGQSVSVATLLMTICFVFMWIFSGLLVNLTTIASWSLQVFSIP	574
543	Db	MTSLVAQSGLLIGAASTLSQVATFVGPVTAIPVLLFSGFVSFDTIPTYLOMYSIYV	602
575	QY	RYGFT-----ALQHNFEIQLNFCPGINATGNNPCYATCTGEYLVKIGDLSPWGLW	627
603	Db	RYGFEQVILSYGLDRED-----LHCDIDTCHQK---SEALIRE-LDVENAKLY	649
628	QY	KXHALACMIVFLTIAYLKULF	650
650	Db	LDPIVLGHFFISLRLIAYFLVRY	672

RESULT 12	
ABG5 MOUSE	
ID--ABG5 MOUSE	STANDARD;
	PRT; 652 AA.

OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus  
OX NCBI TaxId:10090;

RC STRAIN=C57BL/6; TISSUE=Liver;  
RX MEDLINE=20578753; PubMed=11138003;

Query Match 21.1%; Score 706.5; DB 1; Length 678;

3 SSNTEVFIPVSGQNTNGEPATASNDL---KAPT-----EGAV-LSFHNICVRVKLKGFLP 54

38 SSNWEA---TETDLLNGHLKKVDNNL TEAQRFFSSLPRAAVNIEFRDLSYSVEGPWW--92

114 PAN-FKCN SGYVQDDVVGTLTVRENQFSAALRLATMTNHEKNERINRVIQELGLDK 172

THE UNIVERSITY OF CHICAGO PRESS





```
FT TRANSMEM 478 498 POTENTIAL.
SQ SEQUENCE 598 AA; 66906 MW; 9D6414E06898E343 CRC64;
Query Match 20.5%; Score 686.5; DB 1; Length 598;
Best Local Similarity 30.3%; Pred. No. 1.2e-39;
Matches 186; Conservative 124; Mismatches 261; Indels 43; Gaps 14;
QY 61 KEILNININGKPG-LNAILGPTGGGKSSLLDVLAARKDPS-GLSGDVLINGAPRPAN-- 116
DQ 7 KEILNININGKPG-LNAILGPTGGGKSSLLDVLAARKDPS-GLSGDVLINGAPRPAN-- 116
DQ 7 KEILNININGKPG-LNAILGPTGGGKSSLLDVLAARKDPS-GLSGDVLINGAPRPAN-- 116
QY 117 -PKNSGVYQDVVVGTLTVRENIQFSAALRLATT-MTNHEKNERINRVTOELGLDKVA 174
DQ 55 KIRENSAFVQDHMFVGTAREHLQFVARLMRGDQYVSDHERQLRVEQVLTQWGLKCA 124
QY 175 DSKVG-TQPIRVSGGGRKRTSIGMELTDSIILDEPTTGDSSTANAVLLILKMSK 233
DQ 125 DTVIGIPNQLKGLSGCKRLSFASEILTCPRKILFCDEPTSGDLAFWAGHVQALRSLAD 184
QY 234 QGRTIIFSHQPRYSIFKLFDSLTLLASGRLMFHPGPAQALGYPESAGYHCEAYNPADF 293
DQ 185 NGMTVIIITHQSSHVYSLFNNVCLMAGRVILYLGQDQVPLFEKCGYPCPAYNPADH 244
QY 294 F--LDIINGDSTAVALARNEEDFKATEIIEPSKQDKPLIEKLAETIYVNSPYKTKA--- 347
DQ 245 LIITLAVIDSD-----RATSM-----KTIKIROGFLSTDLGQSVLAIGN 284
QY 348 --ELHQLS--GGEKKKTIYFKEISYITSPCHLRWYSKSPKXLLGNPQASIAQIIVT 402
DQ 285 ANKLRAAEFVTSQDSISEKTFNFQDYNASFWTQFLAFWRSWLTIVIRDPNLLSVRLIQI 344
QY 403 VVLGVIGIAYFGLKNDSTGIQNRAGVLFLLTNQCF-SVSARELVFVEKKLFTHEIYIS 461
DQ 345 LITAFITGVTFVTPVTPATIIISINGINFMFNRMFMQLQPNVPVITAEPLVIRENAN 404
QY 462 GYRVSSVFLGLSLLDPLMTWLPSTIFCIYVFMGLKPKADAFVFMVFLMWVAYSAS 521
DQ 405 GYRTSAYFLAKNIAELPOYIILP-ILYNTIVYMSGLYFNFNFCFASLVTILITNVAI 463
QY 522 SMALATAAGQSVVATLMTICFVFMIFSLVNLTTIASWLSWLGQYFIPRYGFTAL 581
DQ 464 SISAVATIFANTDVAMTILPIFVVPIMAFGFFITFADIPSYFKWLSLSVFKYGEAL 523
QY 582 QHNEFLGQNFCLNATGNNPCNYATCTGEEVLVKGQDLSWGLWKNHVALACMIVFL 641
DQ 524 AINWDSIKVIECFNNSMTAFALDSCPKNGHQVLESIDFSASHKIFDISILFGMFIGIR 583
QY 642 TIAYKLFLFKYYS 655
DQ 584 IIAVYVALL-IRSYN 596
RESULT 14
ABGS_HUMAN STANDARD; PRT; 651 AA.
AC QSH22;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE ATP-binding cassette, sub-family G, member 5 (Sterolin-1).
GN ABCG5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANT GLU-604.
RC TISSUE=Liver;
RX MEDLINE=20553648; PubMed=11099417;
RA Berge K.E., Tian H., Graf G.A., Yu L., Grishin N.V., Schultz J.,
RA Kwockovich P., Shan B., Barnes R., Hobbs H.H.;
RT "Accumulation of dietary cholesterol in sitosterolemia caused by
RT mutations in adjacent ABC transporters.";
```

DR PROSITE; PS00893; ABC\_TRANSPORTER\_2; 1.  
 KW ATP-binding; Glycoprotein; Transmembrane; Transport; Polymorphism;  
 KW Disease mutation.  
 FT DOMAIN 1 383 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 384 404 1 (POTENTIAL).  
 FT DOMAIN 405 421 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 422 442 2 (POTENTIAL).  
 FT DOMAIN 443 462 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 463 483 3 (POTENTIAL).  
 FT DOMAIN 484 503 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 504 524 4 (POTENTIAL).  
 FT DOMAIN 525 549 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 550 623 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 624 644 6 (POTENTIAL).  
 FT TRANSMEM 645 651 CYTOPLASMIC (POTENTIAL).  
 FT NP\_BIND 86 93 ATP (POTENTIAL).  
 FT CARBOHYD 584 584 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 591 591 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARIANT 146 146 E -> Q (in sitosterolemia).  
 FT VARIANT 389 389 /FTid=VAR\_012244.  
 FT VARIANT 419 419 R -> H (in sitosterolemia).  
 FT VARIANT 419 419 /FTid=VAR\_012245.  
 FT VARIANT 419 419 R -> H (in sitosterolemia).  
 FT VARIANT 419 419 /FTid=VAR\_012246.  
 FT VARIANT 550 550 R -> P (in sitosterolemia).  
 FT VARIANT 550 550 /FTid=VAR\_012247.  
 FT VARIANT 604 604 R -> S (in sitosterolemia).  
 FT VARIANT 604 604 /FTid=VAR\_012248.  
 FT SEQUENCE 651 AA; 72503 MW; 950BABCBB6A1536 CRC64;  
 Query Match 20.4%; Score 682.5; DB 1; Length 651;  
 Best Local Similarity 29.2%; Pred. No. 2.5e-39;  
 Matches 182; Conservative 138; Mismatches 249; Indels 55; Gaps 18;  
 QY 13 SQNTWGFPTASNDLKAFTEGAVLSFHNICRVKLSGFLPCRKVEKEILSNINGIMK 72  
 DB 21 SQSLEGPATAP---EPHSLGLTHASVSVHRVPMWDITSCROQWTQKILKDSLYVE 77  
 QY 73 PG-LNALIGTGKGSLLDVLAAKDPG-LSGDVLING-APRPANFKCNSGYVVQDDV 129  
 DB 78 SGQIMCLGSSGSKITLLDAMSGRLGRAGTFLGEVYVNGRALRRQFQDCSYVLQSDT 137  
 QY 130 VMGTLVRENLPQSAALRLATTWNHEKNERINRVQELGLKVDKSVGTQFIRGVSQ 189  
 DB 138 LLSLTVRETHYTLALLAIRGNPG-SFOKKVEAVNAELSLSHVADRLIGNYSLOGISTG 196  
 QY 190 ERKRTSIGMELITDPSLFLDEPTTGLDSTANAVLLKRMKSGQRTIIFSTHOPRYSI 249  
 DB 197 ERRVSIAGLLQDPKVMLEDEPTTGLDCMTANQIVVLLVLAARNRIVVLIHQPSRL 256  
 QY 250 FKLFDSTLLASGRLMFHPQAQALGYFESAGVHCAYNNPADFFDIIDINGSTAVALNR 309  
 DB 257 FQLFDKIALISFGLIFPCGTPAEMLDFFNDGCVPCPSHSNPFYMDLTSVDTQ----SK 312  
 QY 310 EEDFKATEIIEPSKODKFLKELAEIYVNSSFYKAKAELHQLSGGEKKKKTIVFKESY 369  
 DB 313 ERE-----LETGR-----VOMTESAYKKAICHT-----LKNIERMKHKLTPMVPF 356  
 QY 370 TT----SFCQLRWKSRSPKLLGNPQASIAQIIVTVVLGLVIGAIYFGLKNDST---- 421  
 DB 357 KTKDSPGVFSKGLGVLLRVRVRLNRKLAIVITRLQLNGLFL--LFFVLRVRSNVLKG 414  
 QY 422 GIONRAGVLP-FLTTWQCFSSVAVELFVVEKLFTHIYISGYRVSSYFGLKLSLDLAP 480  
 DB 415 AIQDRVGLLYQFVATPYTGMLNAVLFPVIRAVSDQESQDGLYQKQOMLVAL-HVLP 473  
 QY 481 MTLVLSIIFTCIVYFVGLKPKADAFVMMFTLM--MVAYSASSMALAIAAGQSVVSA 537  
 DB 474 FSVVATMIFSSVCVWTLLGLHPEVARFGYFSAALLAPHLIGEFTLLVLLGIVQNPENIVNSV 533

QY 538 TLLMTICFVFMWIFSGLLVNLTTIASWLSLQVIESIPRYGVTALQHNERLQNFQPCG--- 594  
 DB 534 VALLSIAGV--LVGSGFLRNQMPFPKLIISIFTFQKICSEILVNVNFGVGLNFTGSSN 591  
 QY 595 LNATGNPNPCNYA-----TCTG 610  
 DB 592 VSVTNPMACTQIGIQIENKTCG 615  
 RESULT 15  
 'ABGS\_RAT  
 ID ABGS\_RAT STANDARD; PRT; 652 AA.  
 AC Q99PE7; Q8C1Q4;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE ATP-binding cassette, sub-family G, member 5 (Sterolin-1).  
 GN ABCG5.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Small intestine;  
 RX MEDLINE=20578753; PubMed=11138003;  
 RA Lee M.-H., Lu K., Hazard S., Yu H., Shulenin S., Hidaka H., Kojima H.,  
 RA Allikmets R., Sakuma N., Pegoraro R., Srivastava A.K., Salen G.,  
 RA Dean M., Patel S.B.;  
 RA "Identification of a gene, ABCG5, important in the regulation of  
 RT dietary cholesterol absorption.";  
 RL Nat. Genet. 27:79-83 (2001).  
 RN [2]  
 RP REVISION TO 2.  
 RP Lu K., Lee M.-H., Patel S.B.;  
 RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A. TISSUE SPECIFICITY, AND VARIANT CYS-583.  
 RC STRAIN=GH, SHR, SHRSP, Sprague-Dawley, Wistar, Wistar Kyoto, and WKA;  
 RX PubMed=12783625;  
 RA Yu H., Pandit B., Klett E., Lee M.H., Lu K., Helou K., Ikeda I.,  
 RA Egashira N., Sato M., Klein R., Batta A., Salen G., Patel S.B.;  
 RT "The rat STSL locus: characterization, chromosomal assignment, and  
 RT genetic variations in sitosterolemic hypertensive rats.";  
 EL BMC Cardiovasc. Disord 3:4-4(2003)  
 CC -!- FUNCTION: Transporter that appears to play an indispensable role  
 CC of the selective transport of the dietary cholesterol in and out  
 CC of the enterocytes and in the selective sterol excretion by the  
 CC liver into bile.  
 CC -!- SUBUNIT: May form heterodimers with ABCG8 or be tightly coupled to  
 CC ABCG8 along a pathway regulating dietary-sterol absorption and  
 CC excretion (by similarity).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
 CC -!- TISSUE SPECIFICITY: Expressed only in liver and intestine.  
 CC -!- POLYMORPHISM: The polymorphism at position 583 is found in strains  
 CC SHR, SHRSP and Wistar Kyoto which are both hypertensive and  
 CC sitosterolemic. Strains which are hypertensive but not  
 CC sitosterolemic do not contain a polymorphism at this position.  
 CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)  
 CC subfamily.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 CC EMBL; AF312714; AAC53098.3; -;  
 CC EMBL; AY145899; AAN64275.1; -;  
 CC InterPro; IPR003593; AAA\_ATPase.  
 CC InterPro; IPR003439; ABC\_transporter.

DR Pfam; PF00005; ABC tran; 1.  
DR ProDom; PD000006; ABC transporter; 1.  
DR SMART; SMO0382; AAA; 1.  
DR PROSITE; PS00211; ABC TRANSPORTER 1; 1.  
DR PROSITE; PS00893; ABC TRANSPORTER 2; 1.  
KW ATP-binding; Glycoprotein; Transmembrane; Transport; Polymorphism.  
FT DOMAIN 1 385  
FT TRANSMEM 386 406  
FT DOMAIN 407 422  
FT TRANSMEM 423 443  
FT DOMAIN 444 463  
FT TRANSMEM 464 484  
FT DOMAIN 485 504  
FT TRANSMEM 505 525  
FT DOMAIN 526 529  
FT TRANSMEM 530 550  
FT DOMAIN 551 624  
FT TRANSMEM 625 645  
FT DOMAIN 646 652  
FT NP BIND 87 94  
FT CARBOHYD 585 585  
FT CARBOHYD 592 592  
FT VARIANT 583 583  
FT SEQUENCE 652 AA; 73372 MW; 49FEF7372269299D CRC64;  
SQ  
Query Match 20.4%; Score 682.5; DB 1; Length 652;  
Best Local Similarity 28.9%; Pred. No. 2.5e-39;  
Matches 185; Conservative 142; Mismatches 237; Indels 77; Gaps 22;  
QY 2 SSSNVEVPIVSGQNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLKSG-----FLPCRK 57  
DB 21 SQSSLE-----EGSVTGEARHS-----LGVINVSFVSQRVRGPMWNKSCQQ 63  
QY 58 PVEKEILNINIMKPGLN-ATLPTGGKGLLDVLAARKDPGQ-LSGDVLINGAP-RP 114  
DB 64 KWDKILKDVSLYIESGQTMCLGSSGSGKTLDDAISGRLLRRTCTLEGEVFNVCGLR 123  
QY 115 ANFKCNMGVQDDVVMGTIVRENQPSAALRLATTNHNKNERINRVIOELGLDKVA 174  
DB 124 DQFQDCVSLQSDVFLSGLVRETRYTAMALRSSADF-YDKKVEAVLTSLSHVA 182  
QY 175 DSKVGTQPIRGVSGGERKRTSIGMELITDPSILPLDEPTTGLDSTANAVLLLLKRMKQ 234  
DB 183 DQMGNYNFGGSSGERRRVSIAQLQDPKVMMLDEFTTGLDCMTANHIVILLVARR 242  
QY 235 GRTIIFSIHQPRYSIFKLPDSLTLLASGRLMFHGPAQALGYFESAGVHCEAYNPADFF 294  
DB 243 NRIVIVTIHQPRASELFHFHFDKIALITYGELVFCGTPPEMLGPFNNCGYPCPEHSNPFDFY 302  
QY 295 LDIINGDSTAVALNRE-EDFKATEIIEPSKQDKPLIEKLAIEIYVNSSPYKETKAEHLHLS 353  
DB 303 MDLTSVDIQ--SREHEIETKRVOMLESARFROSDICHKI-----LE 341  
QY 354 GGEKKKKTIVFKIISYTTSS-----FCHLRWVSKSFKNLGNPOASIAQIVTVVLGLV 408  
DB 342 NIERTRHKLTPMPVPPKTKNPPGMFC-KLGVLLRRVTRNLRNKQVIMRLVQNLGLF 400  
QY 409 IGAIVFGLKNDST----GQNEAGVLFELTTNQCFSS-VSAVELFVVEKLFIEHYISGY 463  
DB 401 L--IFVLLRVQNMMLKGAQVQGLLYQLVGATPTGMLNAVNLFPMLRAVSDQSDGL 458  
QY 464 YRVSSYFLGLKLSDLPLMTMLPSIIFTICIVFMLGLPKPADAFFVMMFTLM---MVAYSA 520  
DB 459 YQKWOMLLAVL-HALPFSIVATVIFSSVCYWTGLGYPEVARFGYFSAALLAPHLIGEFL 517  
QY 521 SSMALAIAGQSVSVATILMICVFPMIFSGLLVNLTTIASWLSWLYQFSIPRYGFTA 580  
DB 518 TLVLLGMVQNPINVSIVALLSIS--GLLIGSGFIRNIEEMPILKILGYFTFKYICEI 575  
QY 581 LQHNEFLQGNF-CFGLN-ATGNNP-CNYA-----TCTG 610  
DB 576 LVVNEFYGLNFTCGGNTSVNPNPNCMTQGIQFIEKTCPG 616

Search completed: March 23, 2004, 18:30:03  
Job time : 20 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

QM protein - protein search, using sw model  
Run on: March 23, 2004, 18:27:43 ; Search time 20 Seconds  
(without alignments)  
3150.272 Million cell updates/sec

Title: US-09-961-086-1  
Perfect score: 3352  
Sequence: 1 MSSSNVEVFPVSGQNTNGF.....MIVFLTIAYLKLLFLKKYS 655

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues  
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78.\*  
1: PIR1.\*  
2: PIR2.\*  
3: PIR3.\*  
4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2849.5	85.0	656	JC7860	brain multidrug re
2	835.5	24.9	1049	S19421	ATP-dependent perm
3	812	24.2	687	FFFW	white protein - fr
4	800.5	23.9	737	T46101	ABC transporter-11
5	774	23.1	646	C86441	probable ABC trans
6	767.5	22.9	687	D96553	hypothetical prote
7	746	22.3	649	A84509	probable ABC trans
8	745	22.2	725	T47652	ABC transporter-11
9	741.5	22.1	739	T45891	ABC transporter-11
10	739	22.0	678	H96552	hypothetical prote
11	726.5	21.7	708	T47650	ABC transporter-11
12	721	21.5	635	T08934	hypothetical prote
13	720.5	21.5	755	C84791	probable ABC trans
14	717.5	21.4	740	T02567	probable ATP-bindi
15	712.5	21.3	609	E96742	probable ABC trans
16	710	21.2	638	G02068	white homolog - hu
17	708	21.1	646	JC7777	ATP binding casset
18	707.5	21.1	547	T31543	hypothetical prote
19	707.5	21.1	720	T47648	ABC transporter-11
20	677	20.2	559	B98474	protein C05D10.3 [
21	676	20.2	725	C94423	probable ABC trans
22	669	20.0	659	E96313	hypothetical prote
23	664	19.8	608	T34391	hypothetical prote
24	660	19.7	662	T47649	ABC transporter-11
25	658.5	19.6	1294	T27690	probable membrane
26	657.5	19.6	1450	T45898	ABC transporter-11
27	656.5	19.6	590	T24588	protein F12M16.17
28	649.5	19.4	633	T19189	hypothetical prote
29	645.5	19.3	577	T04229	ABC-type transport

30	639	19.1	658	2	T31958	hypothetical prote
31	636.5	19.0	639	2	G88839	protein C10C6.5 [i
32	636.5	19.0	695	2	T21109	hypothetical prote
33	627.5	18.7	610	2	T19333	hypothetical prote
34	612.5	18.3	1501	2	S50992	SNQ2 protein - yea
35	612	18.3	1530	2	S52239	brefeldin A resist
36	603	18.0	1530	2	T52010	hypothetical prote
37	602	18.0	1511	2	A53151	pleiotropic drug r
38	594	17.7	1564	2	S55517	probable transport
39	593	17.7	1443	2	T02431	probable ABC trans
40	590	17.6	1469	2	H96622	probable ABC trans
41	585	17.5	1333	2	S63403	probable membrane
42	580.5	17.3	1420	2	T02644	ABC-type transport
43	580.5	17.3	1529	2	S69688	hypothetical prote
44	572	17.1	1413	2	G84790	probable ABC trans
45	564	16.8	1466	2	T30566	ATP-binding casset

ALIGNMENTS

RESULT 1

JC7860  
brain multidrug resistance protein, BMDP - pig  
C:Species: Sus scrofa domestica (domestic pig)  
C:Date: 18-Nov-2002 #sequence\_revision 18-Nov-2002 #text\_change 31-Mar-2003  
C:Accession: JC7860  
R:Eisenblatter, T.; Galla, H.J.  
Biochem. Biophys. Res. Commun. 293, 1273-1278, 2002  
A:Title: A new multidrug resistance protein at the blood-brain barrier.  
A:Reference number: JC7860; MUID:22050127; PMID:12054514  
A:Accession: JC7860  
A:Molecule type: mRNA  
A:Residues: 1-656 <EIS>  
A:Cross-references: GB:AJ420927  
A:Experimental source: brain  
C:Comment: This protein, a new transport protein of the ATP-binding cassette (ABC) sup  
exclusion of xenobiotics from the brain and participates in drug transport across the  
C:Genetics:  
A:Gene: bmdp

Query Match	85.0%	Score	2849.5	DB 2	Length	656			
Best Local Similarity	84.3%	Pred. No.	4.9e-195						
Matches	553	Conservative	44	Mismatches	58	Indels	1	Gaps	1
QY	1	MSSSNVEVFPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCR	PVE	60					
Db	1	MSSSNYQVSIPLMSKNTNGLPGSSSNEILKTSAGGAVLSFHDICRVKVSGLFCRK	IVE	60					
QY	61	KEILSNINGIMKPGLNAILGPTGGKSSLLDVLAAKDPGSLSGDVLINGAPRANFKCN	120						
Db	61	KEILTNINGIMKPGLNAILGPTGGKSSLLDVLAAKDPHGLSGDVLINGAPRANFKCN	120						
QY	121	SGYVQDDVVMGTLTVRENLOFSAALRLATMTNHEKNERINRVIOELGLDKVADSKVGT	180						
Db	121	SGYVQDDVVMGTLTVRENLOFSAALRLPTMTNHEKNERINRVIOELGLDKVADSKVGT	180						
QY	181	QPIRGVSGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVLLILKRMKSQKQRTIIF	240						
Db	181	QPIRGVSGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVLLILKRMKSQKQRTIIF	240						
QY	241	SIHQPRYSIFKLFDLSITLLAGRLMFHGPQAELGYFESAGYHCEAYNNADFFLDING	300						
Db	241	SIHQPRYSIFKLFDLSITLLAGRLMFHGPAREALGYFASIGYCNCPYNNADFFLDVING	300						
QY	301	DSTAVALNR-REDFKATEIIEPSKODKPLIEKLABIYVNSSFYKETAELHQLSGEKKK	359						
Db	301	DSAAVULSRADDEGAQEPPEPEKDTPLIDKLAIFYNSSFDDTKVELQDFSGRKKK	360						
QY	360	KITVFKEISYTFCHQLRWWSKRSFKNLLGNPQASIAQIIVTVLGLVIGAIYFGLKND	419						
Db	361	KGSVYKRYVYTTFSCHQLRWWSKRSFKNLLGNPQASVAQIIVTILGLVIGAIYFDLND	420						

[illegible]



Query Match	23.9%	Score 800.5	DB 2	Length 737
Best Local Similarity	31.4%	Prod. No. S.6e-49		
Matches 211	Conservative 132	Mismatches 227	Indels 101	Gaps 21
QY	27	DLKAFTEGAV-----LGFHNICVYRVKLKSGFLPCRPKVPKEILSNINGIMKFG-	74	
Db	124	DIEAATSVVKFOAEPFFIYLAFIDITYKVTKG---MTSSSEKSLINGISGSAYDGE	179	
QY	75	LNALGPTGGKSSLLDLVAARKDPGSLGSDVLINGAPRANPKNSGVYVDDVYMGTL	134	
Db	180	LLALMGPSGGKTTLLNALGGRFNQNIIGSVSYNDKPYSKLKTTRIIGFTVDDVLFPHL	239	
QY	135	TVRENLOFSAALRLATTMTNHEKNERINRVIOELGLDKVADSKVGTQFIRGVSGGERKRT	194	
Db	240	TVKETLYIYALLRLPKTLTQEKEQRAASVIOELGLERCQDTWIGSGFVRGVSGGERKV	299	
QY	195	SIGMBELTDSPIELDFPTGILDSSTANAVLLLLKMSKQGRITIPSHQPRYSIFKPLD	254	
Db	300	CIGNEIMTNPSSLLDPTSSLDSTALKIVQMLHCIAKAGKITVTTIHQPSRSLRHFHFD	359	
QY	255	SLTLLASGRLMFHGPAOEALGYFESAGYHCEAYNNPADFFLOIINGDSTAVALNREEDPK	314	
Db	360	KLIVLSRGSLLYTFGKASEANSYFSSIGCSPLLAMNPAEFLDLVNGNMNDIS----	411	
QY	315	ATEIIEPSKQDKPIUEKLABIYVNS-----SFYKETHAEHLHQLS-----GGE	356	
Db	412	----VPSALKEKMKIIRL-BLYVRNVKCDVETOYLEAYKTQIAYNMEKMLMAPVLPDDE	466	
QY	357	KKKIKTTPKE--ISYTTSPC-HQLRWVSKRS---FNKLLGNPQASIAQIITVVVLGLVI	409	
Db	467	VKLMTICPKRWGSLSWEQYCLSLRGIKERRHDYFSWL-----RVIQVLSTAI--IL	517	
QY	410	GAIVFGLKNDSTGIO-NRAGVLFFLTNTNQCFSSV-SAVELFVVVEKKLFTHEIVISGYRV	467	
Db	518	GLLWV--QSDITSORPTRSGLLFFIAVFWGFFVFVTAITFPQERAMLKSERESNNYRLS	575	
QY	468	SYFLGKLLSDLLMTMLPSIIFTCIVYFMLGLPKPADAFVMMFTLMWVAYSASSMALAI	527	
Db	576	AYFVARITSD-LPDLILPLVLVYVYFMAGLRLRAESFFLSVLTFLVCIVAAQGLGLAI	634	
QY	528	AAGQSVSVATLLMTICFVFMIMPISGLLVNLTITASLWSLQVFSIPRYVGFTAL---	584	
Db	635	GASLMDLKKATTLASVTWTFMLAGGYFVK--KVPPFIAIRFMSFNVTYKLLVKVQVE	692	
QY	585	EFLGQNFQCGLNATGNPCNVATCTGBEYVLVKGQIDLSPWGLMKNHVALACMIVFLTTIA	644	
Db	693	ELW-----ESVNGEE--IESGL-----KEYSALVAMLIIGVRLVA	724	
QY	645	YKLKFLFLKKYS	655	
Db	725	YPSLRMRMKLHS	735	

RESULT 5  
C86441  
probable

## RESULT 6

583 --AGEPKMTGEQVINKIFGVQVTHSKWWDLSAIVLILVCYRILFFIVLKLK 631

## RESULT 7

A84509  
Probable ABC transporter [imported] - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
C/Accession: A84509  
M.; Koo, H.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Lin, X.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umavam, L.; Tallon, L.;  
euss, D.; Nierman, W.C.; White, O.; Bisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A/Reference number: A84420; MUID:20083487; PMID:10617197  
A/Accession: A84509  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-649 <SPO>  
A/Cross-references: GB:AEO02093; NID:94558665; PID:AAD22683.1; GSPDB:GN00139  
C/Genetics:  
A/Gene: At2g13610  
A/Map position: 2  
C/Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1; ATP-B

Query Match 22.3%; Score 746; DB 2; Length 649;  
Best Local Similarity 33.5%; Pred. No. 3.6e-45;  
Matches 203; Conservative 116; Mismatches 239; Indels 48; Gaps 18;  
QY 61 KEILSNIINGIMKP-GLNAILGPTGGKSSLLDVLAAARKOPSGDVLINGAP-RPANFK 118  
DB 60 KHVLLGVTCKAPWEILAIIVGSPGAGKSSLEILAAARLIPO--TGSVYVYKRPVDRANFK 117  
QY 119 CNSGVVQDDVMGVTITVRENLOFSAALRLATMTNHEKNERINRYIQELGLDQVADSKV 178  
DB 118 KISGVYTKDTLPFLIVESTLTFSAKRL--KLPADELRSRVKSLVHELGLAEVATARV 175  
QY 179 GTQFIRGVGGERKRTSIGMELTDPISILFDEPTTGLDSSSTANAVALLKRMK-QGR 237  
DB 176 GDDSVRGISGGERRRVSGIVEIHDPKVILDEPTSGLDSTALLIDMLKMAETRGRT 235  
QY 238 IIFSIHOPRYSIKLPDLSLLASGLMFGPAQEAALGFESAGYHCEAYNNPADFLDI 297  
DB 236 ILTIHOPGRIVKQFNQSVILLANGSTLKQGSVDQLGVLRSNGLPHHPHENIVEFAIES 295  
QY 298 INGDSATAVLRNEDFATBIIEPSKQDKPLIEKLAEIYVNSSFYKETKAEHLQLSGGRK 357  
DB 296 I--ESITKQQLQESRRAAHVLP--QTTLOEKRSSEDSQGES--KSGKFTLLQLFQQR 348  
QY 358 KKKI-TVFKELISYTSFCH-----QLRWVKSFPKULLGNPQASIAQIIVTVLGLVIGAI 412  
DB 349 VADVTGNVIAETFRDPFANSRLBETMILTRFKNIFRTKELFACRTVOMLGGIVLGLI 408  
QY 413 YFGKNDSTGIONRAGVLFLLTNQCPSSVSVAVELFWVEKKLFIHEYISGYRVSSVYFLG 472  
DB 409 FHNLDKLGAREVGLFAILFTLLSTIEALPILOREILMKETSSGSRVSSYAVA 468  
QY 473 KLSDDLPMWLPISIIITCIYVFMGLKPKADAFVMMFTLMVAVSASSMALAIAA--- 529  
DB 469 NGLV-YLFFLLILAILESTPYVWLVGLNPSMAFLHFSLLILILXTANSVVVCFALVP 527  
QY 530 ---GQSVSVATLMTICFVFMIFSGLLVNLTTIASMLSLQYPSIPRYGTALQHNE 585  
DB 528 NFIQNSVIGS---VMSGFP-----LFGSYFISNHEIPGYWIFWHYISLFKYPEGFLINE 580  
QY 586 FLGNFCPLGNATNPNCAVTCATGEBYLVKQIGIDLSPWG---LWNKHVALACMIVIFLT 642  
DB 581 FSKSNKC---LEYGFGKC---LVTEDDLKE-----BRYEESRWNRVNVIMLCFVLLYRF 629  
QY 643 IAYLKL 648  
DB 530 ISYVIL 635

## RESULT 9

T45891  
ABC transporter-like protein - Arabidopsis thaliana

## RESULT 8

T47652  
ABC transporter-like protein - Arabidopsis thaliana  
N/Alternate names: protein T26112.10  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 19-May-2000  
C/Accession: T47652  
R.; Monfort, A.; Casacuberta, E.; Puigdomenech, P.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.  
submitted to the Protein Sequence Database, February 2000  
A/Reference number: 224471  
A/Accession: T47652  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-725 <MON>  
A/Cross-references: EMBL:AL132954  
A/Experimental source: cultivar Columbia; BAC clone T26112  
C/Genetics:  
A/Map position: 3  
A/Note: T26112.10  
C/Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1; ATP

Query Match 22.2%; Score 745; DB 2; Length 725;  
Best Local Similarity 29.3%; Pred. No. 5e-45;  
Matches 193; Conservative 130; Mismatches 274; Indels 62; Gaps 13;  
QY 36 VLSFHNICRYVKLKGFLPCRKPVEKEILSNIINGIMKPG-LNAILGPTGGKSSLLDVL 94  
DB 72 VLNFNQYDVTLLRRRFGFSRQNGVKTLLDDVSGEASDGLILAVLGASGAKSTLIDALA 131  
QY 95 ARKDPGSLGSDVLINGAP--RPANFKNSGVVQDDVMGTLTVRENLOFSAALRLATMT 152  
DB 132 GRVAEGLRGSVTLNGEKVLQSRLLKVIISAYVMODLLFPMLTVKETIMFASERLPRSL 191  
QY 153 TNHEKVERINRVIQELGLDQVADSKVGTQFIRGVGGERKRTSIGMELTDPISILFDEP 212  
DB 192 SKSKQWEREALIDQLGLNANAVTIGDEHGVSGGERRRVSGIILHDPVILFDEP 251  
QY 213 TTGLDSTANAVALLKRMKSGRTIIFSIHOPRYSIFKLPDLSLLASGLMFGPAQEA 272  
DB 252 TSGLDSTNFAVWVQVVKRIAGSGSIVIMSHOPSARIVELIDRLIILSRGKSVENGSPAS 311  
QY 273 ALGYFESAGYHCEAYNNPADFLDI-----NGDSTAVLRNEDFATBIIEPSK 322  
DB 312 LFGFSDFGRIPEKENISEFALDOLRELESGNGETKALVDNFNEKQONKISLQSPQT 371  
QY 323 ---QDKKPLIEKLAEIYVNSSFYKETKAEHLQLSGGRKKTIVFKELISYTSFCHQRLW 379  
DB 372 NKLDQDRSLSLKEA---INASV---VSGSRKSNPTSMETVSSYANPSLFEFTI 423  
QY 380 VKRSFKULLGNPQASIAQIIVTVLGLVIGAIYFGLKNDSTGIONRAGVLFLLTNQCF 439  
DB 424 LAKYMKQWIRMPBELVGTATVMTVTCGLATVYVKLDHTPRGAQERTLFAFVVPNIFY 483  
QY 440 SSVSAVELFVVEKKLFIHEYISGYRVSSVYFLGKLLDPLMTLPSIIFCIYVFMGL 499  
DB 484 CCLDNVPFIOERYIFLRETHNAYRTSSVYISHLSV--LPQLLAPSLVFSALFTWTVGL 542  
QY 500 KPAKADAFVMMFTLMVAVSASSMALAIAAQSVSVATLMTICFVPMFTPSGLLVNLT 559  
DB 543 SGGLEGFVYCLLIYASFWSGSVVTFISGVVPMILCYMVSIYLYACLLSGIFYNRD 602  
QY 560 TIASVLSWLOQPSIPRYGTALQHNEFLQNFQ-----PGLNATG----- 599  
DB 603 RIFPFWTFHYSILKPYEAVLINEPDDPSRCFVRGVQVFDSTLLGVSQSGVKVLET 662  
QY 600 -----NPNCAVTC--TGEYLVKQIGIDLSPWG---LWNKHVALACMIVIFLT 650  
DB 663 LSKSLRTRKITECTLRGTGSDLLAQOIGTQLSKWD-----CLWITFASGLFFRIILF 712

N:Alternate names: protein F4P12.210  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Mar-2000  
C:Accession: T45891  
R:Blöcker, H.; Meves, H.W.; Lencke, K.; Mayer, K.F.X.; Quétier, F.; Salanoubat, M.  
submitted to the Protein Sequence Database, January 2000  
A:Reference number: Z33016

A:Accession: T45891  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-739 <BLO>  
A:Cross-references: EMBL,AL132966  
A:Experimental source: cultivar Columbia; BAC clone F4P12  
C:Genetics:  
A:Map position: 3  
A:Note: F4P12.210  
C:Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1; ATP-B

Query Match 22.1%; Score 741.5; DB 2; Length 739;  
Best Local Similarity 28.3%; Pred. No. 9,1e-45;  
Matches 193; Conservative 123; Mismatches 256; Indels 111; Gaps 17;  
QY 36 VLSFNNICRYVKLKGK--LPC--RKPE-----KEILSNINGIMKPG-LNAIGLPT 82  
DB 87 VLSFKDLTYSVKIKKKKPPCCGNSPPDGNMNTKVLINGISGEAREGEMAVILGAS 146  
QY 83 CGKSSLLDLVAARKDPSGLSDVLINGAPRANF-KCNSGYVQDDVVMGTLTVRENLQ 141  
DB 147 GSGKTLIDALANRISKESLGDITLNGEVLESSLHVISAIVNQDQLLFPMLIVESTLM 206  
QY 142 FSAALLRLATTMTNHEKNERINRVIOELGLDKVADKVGTOPIRVSGGERKRTSIGMELI 201  
DB 207 FSAEFLPSSLSKKKKARVQALIDQLGLRNAKTVIGDEGHRGVSGGERRVSGITDII 266  
QY 202 TDPSTILFDEPTGTDSTANAVLLLLKXMSKQGTIIFSIHQPRYSFKLFDLSLTLAS 261  
DB 267 HDPIILFDEPTSGUDSTAYVVKVQRIAGSGIVMSIHQPSYRILGDLDLKILFSR 326  
QY 262 GRMFHGPQAQEAALGFESAGHYCEAYNNPADFFLDIINGDSTAVALANREEDFKATEIEP 321  
DB 327 GNTVYSGSPHTLPQFFSEFGHPHPENENKPEALDLIR-----ELED 369  
QY 322 SKQDPLIEKLAELIYNSFYKETAELHOLSGGKKKKIIVFKEI----- 367  
DB 370 PEGTKSLVE-----FHQWBAK--OTSQSRNTNVS LKDAISASIRGKLVS 417  
QY 368 -----SYTTSFCHLRWVSKRFGKLLGNPQASIAQIIVTVVLGVIAGIYFGLKND 419  
DB 418 TNLRSFQTFANPFWTEMLVICKRILNSRRQPELFGIRGLAVLTGMILATIFKLDNS 477  
QY 420 STGIQNRAGVLPFLTNQCFSSVAVELFVVEKKLFIHEYISGYRVSSYFLGKLLSLL 479  
DB 478 PRGIQERLGFAPAMSTTPYTCAEAIPVLOERYIFMETAYNAYRRSSVYLAHTIIS-I 536  
QY 480 PMTMLPSIIFTCIVYFMLGKPKADAFVMMFTLMWVAYSSAMALAIAGQSVSVATL 539  
DB 537 PALIILSAFAASTPSAVLAGSGEGLFPFFFTIILTAFWAGSSFTVFLSGVVSHVMIGFT 596  
QY 540 LMTICFVFMIFSGLLVNTTIASMLSMQYFIPRYGFTALQHNELFGQNF--PGLNA 597  
DB 597 VVVAIILAYELLESFGFISDRIPLYMFWPLYSLVKYPYEVQLQNEFEDPTKCFVRGIOM 656  
QY 598 TGNPNCA-----TC--TGEEYLVKOGI--DLSPWG-LWKNH 630  
DB 657 FNSPLGQVPTAVK;SLKXSMGVLGINVTAETCTVTGIDILKQGGIIEKXNCLW--- 713  
QY 631 VALACMIVFLTIA---YLLKLF 650  
DB 714 -----ITWAGFFRVLV 726

RESULT 10  
H96552

hypothetical protein F5D21.8 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001  
C:Accession: H96552  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, Z.A.; Luros, J.S.; Mailf, R.; Marziani  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: H96552  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-678 <STO>  
A:Cross-references: GB:AE005173; NID:g10092361; PIDN:AAG12770.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: F5D21.8  
A:Map position: 1  
C:Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1; ATP-

Query Match 22.0%; Score 739; DB 2; Length 678;  
Best Local Similarity 30.3%; Pred. No. 1.2e-44;  
Matches 186; Conservative 137; Mismatches 248; Indels 42; Gaps 13;  
QY 61 KEILSNINGIMKPG-LNAIGLPTGGKSLLDVLAARKDPG--LSDGVLLNGAPRAN 116  
DB 28 KRLNGWNGCGPNFILAIMGSGSGKSTLLDALAGRL--AGNVWMSKVLVNGKKRLD 85  
QY 117 FKNSGYVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIOELGLDKVADS 176  
DB 86 FGA-AAVYQEDVLLGLTVRESISYSAHLRLPSKLTREEISDIVEATITDMLGECSDR 144  
QY 177 KVGTQPIRVSGGERKRTSIGMELITDPSILFLDPTGLDSTANAVLLLLKXMSKGR 236  
DB 145 TIGNWHLRGISGEGKRLSALEVLTKPSLLFLDPTGLDSTASAFFVQVLRNLTASGK 204  
QY 237 TIIFSHPQRYIFKLFDSLTLASRLMFMHQAQALGFESAGHYCEAYNNPADFFLD 296  
DB 205 TVVSSHQSGEVFALFDDLLLSGGETVYFGEASAATKFFGEAGFCPPSRNPDSHFLR 264  
QY 297 INGD-----STAVALNR-----EEDFKATEIIEPSKQDKPLIEKLAISIYNSFYKTKAEL 349  
DB 265 CVNSDPDNVTAALVESRRINDSSFSLHQLHETNTLDPL-DDIPTAEINTTLVRKFKCSL 323  
QY 350 HQLSGGCKKKKTVP-----XEISYTSFCHLRWVSKRSPKLLGNPQASIAQIIVTV 403  
DB 324 YAAASRARIQETIASIVGIVTERKKSGSQTNWQLRILTORFINMSRDLGYVWMLIAVYI 383  
QY 404 VLGLVIGALYFGLKNDSTGIQNRAGVLPFLTNQCFSSVAVELFVVEKKLFIHEYISGY 463  
DB 384 VLSICVGSIFFNVRNHTNMVSTAACGFMAGFTMSTGGPQSPTEEMKVFSSRLNGH 443  
QY 464 YRVSSYFLGKLLDLPLMTMLPSIIFTCIVYFMLGKPKADAFVMMFTLMWVAYSSAM 523  
DB 444 YGAVVTVNLSLS-LPFIILMCLSTSSITIVNRFQSGSGSHFFNCLDLCIATIVESC 502  
QY 524 ALAIAAGQSVSVATLLMTIC----FVFMIFSGLLVNTTIASMLSMQYFIPRYGFT 579  
DB 503 MMFIAS-----VVPNFMGLGAGYIGIMVLSAGFFRFPDLPVWVFPVSYNYNGAW 557  
QY 580 ALQ-----HNPFLGQPCFGLNATGNPCNATCTGEEYLVKQGDIDLSPWGLKXNHALACM 636  
DB 558 ALQAYKNEMIGVEY-----DSPPLVPMKKGELILQTVLGINPSSKMLDLAVNM 609  
QY 637 IVIFLTIAVLLKL 649  
DB 610 ILIGYRIAPFAL 622

## RESULT 11

T47650  
ABC transporter-like protein - Arabidopsis thaliana  
N;Alternate names: protein T15C9.110  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 19-May-2000  
C;Accession: T47650  
R;Mewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.  
submitted to the Protein Sequence Database, April 2000  
A;Reference number: Z24470  
A;Accession: T47650  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-708 <NEW>  
A;Cross-references: EMBL:AL132970  
A;Experimental source: cultivar Columbia; BAC clone T15C9  
C;Genetics:  
A;Map position: 3.  
A;Note: T15C9.110  
C;Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F1216.1; ATP-B

Query Match 21.7%; Score 726.5; DB 2; Length 708;  
Best Local Similarity 30.1%; Pred. No. 1e-43;  
Matches 198; Conservative 123; Mismatches 282; Indels 55; Gaps 15;  
QY 36 VLSPHNCYVRKLGFLPCRPVE--KEILSNINGMKPG-LNAILPTGGGKSSLDV 92  
DB 62 LLSFNNLSYNNVLRFRDPSRRTASVKTLLDITGEARDGEILAVLGSGAGKSTLIDA 121  
QY 93 LAARKDPSGLGDLVINGAP--RPANFKCNSGVVQDDVVMGTLTVRENLOFSAALRIAT 150  
DB 122 LAGRAEDSLKGTVLNGEKVQSRLLKVISYVQDDLLFPMVLTKTLPASEFRPLR 181  
QY 151 TMTNHEKNERINRVIQELGLDKVADSKVGTQPIRGVSGGERKRTSIGMELITDPSILFLD 210  
DB 182 SLPSKMKERVETLDQLGLRNAADTVIGDEHGRGVSGGERRVSGIGDIIDHPILLFLD 241  
QY 211 EPTTGLDSTANAVLLLLKMSKQGTITFISHQPRYSIFKFLDSTLLASGLRMFHCPA 270  
DB 242 EPTSGLDSTANFVQVQLKRIAQSGSVIMSIHQPSARIIGLDRLIILSHGKSVFNGSP 301  
QY 271 QEALGYFESAGHYCEAYNNPADFFLDII-----NGDSTAVALNREDFKATEIIEPFSKQD 325  
DB 302 VSLSPFSSFGRIPEKENITEFALDVIRELGSSSEGTDLVEFNEKQCNQTARATQTS 361  
QY 326 KPLIEKLAIIYNNSSPYKETKAEHLQSGEKKKITVPEKISYTSFCHQLRWVSKRSF 385  
DB 362 RVSLKEAIAASV-----SRGKL--VSGSSGANPISMETVSSYANPPLAETFILAKRYI 412  
QY 386 KXLLGNPQASIAQIIVTVVLGIVGAIYFGLKNDSTGIQNRAGVLPFLTTNQCFSVSAV 445  
DB 413 KWIPTFELIGRIGTWVTVGLLATVYWLNTPRGAGRGFFAFGNSTMFYCCADNI 472  
QY 446 ELFWVEKKLFIHEYISGYTRVSGYFYGKLLSDLLPMTLPSIPTCIVTFVFMGLKPKADA 505  
DB 473 PVFIQERYIFLRRTTHNAYRTSSYVISHALVS-LPQLLALSIAFAATFTFWTVGLSGGLES 531  
QY 506 FFWMTFTLMVAYSASSMALAIAAGQSVSVATLMTICFV-FMIFSGLLVNLTTIASW 564  
DB 532 FFYICLIIVAFWSGSSIVTFI-SGLIPNVMMSYMTIAYLYCYLLGGSFYINRDRIPLY 590  
QY 565 LSWLQYFSPRYGFTALQHNFEIQQNFC-----PGLNATG 599  
DB 591 WIWPHVYISLLKYFEYAVLINEFDPSRCFVKGVQVFDGTLTAAEVSHVMVKLLDITLSGSL 650  
QY 600 NNECNVATC--TGEEVLYKGI--DLSPWG-LKKNHVALACMIVIELTIAYLKLFLKK 653  
DB 651 GTKITESTCLRTGPDLLMQQGITQLSKWDCLM---ITLAWGL-FFRILFYLSLLFGSK 704

RESULT 12  
T08934

hypothetical protein F27G19.20 - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 17-Mar-2000  
C;Accession: T08934  
R;Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Banerof;  
submitted to the Protein Sequence Database, May 1999  
A;Reference number: Z16519  
A;Accession: T08934  
A;Molecule type: DNA  
A;Residues: 1-635 <BEV>  
A;Cross-references: EMBL:AL078467; GSPDB:GN00062; ATSP:F27G19.20  
A;Experimental source: cultivar Columbia; BAC clone F27G19  
C;Genetics:  
A;Gene: ATSP:F27G19.20  
A;Map position: 4  
A;Introns: 38/3; 253/1; 304/1; 414/3  
C;Superfamily: fruit fly white protein; ATP-binding cassette homology

Query Match 21.5%; Score 721; DB 2; Length 635;  
Best Local Similarity 30.0%; Pred. No. 2.1e-43;  
Matches 203; Conservative 121; Mismatches 284; Indels 68; Gaps 18;  
QY 1 MSSSNVEFIPVSGQNTNGPPATASNDLKAFTEGAVLSFHNICYRVKLXSG---FLPCRK 57  
DB 1 MDNQEVSMQVETPIAKTNDSDSLPFSIFKKANNPVTLKEENLVYTVKLXDSQCGCFKNDK 60  
QY 58 PVKEILSNINGMKPG-LNAILPTGGGKSSLDVLAAR--KDPGSLGSDVLINGAPRP 114  
DB 61 TEERTIKGLTIGTVKPGCEILAMLGPSGKNTSLTALGGRVGEKGLKNTGNISYNNKPLS 120  
QY 115 ANFKCNSGVVQDDVVMGTLTVRENLOFSAALRLATMTNHEKNERINRVIQELGLDKVA 174  
DB 121 KAVRTGTVQDDALYPNLTVETLVFTALLRLPNSFKKQEKIKAKAVMTELGLDRCK 180  
QY 175 DSKVGTQPIRGVSGGERKRTSIGMELITDPSILFLDPTTGLDSDSTANAVLLLLKMSKQ 234  
DB 181 DTIIGPFLRGVSGGERKRVISIQEILINPSLLFLDPTSGLDSTTAQIRVSIILWELARG 240  
QY 235 GRITIIISHQPRYSIFKFLDSTLLASGLRMFHCPAQEALGVFESAGHYCEAYN-NPADF 293  
DB 241 GRIVTVTIHQF-----SKGNPVYFGLGSNAMDYPASVGYSPFLVERINPSDF 286  
QY 294 FLDIINGDSTAVALNREDFKATEIIEPQKQDKPLIEKLAIIYVNSSFYK-----E 344  
DB 287 LLDIANGKPLLV-----ISCWPSVGSDESQRPEAKAALV---AFYKTNLLDSVINE 335  
QY 345 TKAEHLQSGEKKKITVPEKISYTSFCHQLRWVSKRSFKNLLGNPQA--SIAQIIVT 402  
DB 336 VKQDDDLCKNPRESSRVATNTYGDWPTTWQPCVLLKRLKORRHDSFGMKVAQIF-- 393  
QY 403 VVLGLVIGAIYFGLKNDSTGIQNRAGVLPFLTTNQCFSV-SAVELFVVEKKLFIHEYIS 461  
DB 394 -IVSFLGGLLWQTK--ISRLQDQIGLLFFISSFWAFFPLFQOIPTFPQERAMLKERS 450  
QY 462 GYTRVSGYFYGKLLSDLLPMTLPSIPTCIVTFVFMGLKPKADAFVWMTFTLMVAYSAS 521  
DB 451 GWRLSPYFLSRVVGVD-LPMELILPTCFVLVITYMAGLNHNLANFFVTLLVLLVHVLSG 509  
QY 522 SMALAIAGQSVSVATLMTICFVFMFIPSGLLVNLTTIASWLSWLQVFSIPRYGFTAL 581  
DB 510 GLGLALGALVMDQKSATILGSLVIMLTFLLAGYVYVQHPV--FISMKYVSIQYVYKUL 567  
QY 582 QHNEFLQNTCPGLNATGNN---PCNYATCTGBEYL-VKQIDLSPLWGLWKNHVALACMI 637  
DB 568 ----ILQYITANELYPCGDNGKLRCHVGDPEGIKHGFNSGL-----VSALATAML 615  
QY 638 VIELTIAYLKLFLKK 653  
DB 616 VYRVVIAYIALTRIGK 631

RESULT 13  
G84791

probable ABC transporter [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
C:Accession: G84791  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umavam, L.; Tallon, L.;  
Eisen, J.A.; Salzborg, S.L.; Fraser, C.M.; Venter, C.  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487; PMID:10617197  
A:Accession: G84791  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-755 <STO>  
A:Cross-references: GB:AE002093; NID:94056489; PIDN:AAC98055.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g37360  
A:Map position: 2  
C:Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1; ATP-B

Query Match 21.5%; Score 720.5; DB 2; Length 755;  
Best Local Similarity 28.3%; Pred. No. 2.9e-43;  
Matches 202; Conservative 131; Mismatches 282; Indels 99; Gaps 21;

QY 2 SSSNEVEIPVSCQNTGFPATASNDLKAPTEGA-VLSFHNICVYVKLKGFLP---CRK 57  
DB 63 SSRALGIASPINSA-ASSFNSWASAPASSISSPPVLSFTDLTVSVKIQKFNPLACRR 121  
QY 58 PVE-----KEILSNINGIMKFG-LNALIGTGGKSSLLDVLAAKDPGSLGSDVLNG 110  
DB 122 SGNDSSVNTKLLNGISGEAREGEMVAVIGASGSKSTLDALANRIAKDSLRSITLNG 181  
QY 111 APPRANP-KCNSGYVQDDVVGTTITVRENLFQSAALRLATMTNHEKNERINRVIQELG 169  
DB 182 EYLESMMQKVISAYWQDILLPMTVEETLWFSAEFLPRSLSKKKKARVQALDQLG 241  
QY 170 LDKVADSKVGTGIRGVSGGKRPSIGWELITDPSILFLDEPTGLDSSANAVLLLK 229  
DB 242 LESAAKTVIGDEGHRGSGGRRRYSIGNDIHDPILFLDEPTSGLDSTAYWIKVLQ 301  
QY 230 RMSKQGRITIFSIHOPRYSIFKLPDSTLLASGRMLFHPGAQALGVFPESAGHYCEAYNN 289  
DB 302 RIAQSGSIVMSIHOPSYRIMGLLDQLIFLSKGNVYSGSPHLPPQFSEFKHIPENEN 361  
QY 290 PADFFLDINGDSTAVLNREEDFKATEIIPSKQDKPLIE-----KLAEIYN----- 338  
DB 362 KTEFALDLI-----RELEYS-----TEGKPLVEFHKQWRAKQAPSYNNNNKR 404  
QY 339 ----SSFYKETKAEHLQ---LSCGKXKKITVFEKI-SYTSFCHOLRWKSRGFKNLLG 390  
DB 405 NTNVSLEKATIASISRGKLVSGATNNSSNLTPSFQTFANPFWIEMIVIGKRAILNRR 464  
QY 391 NPQASIAQIIVTVVLGVLGAIYFGLKNDSTGIQNRAGVLFPLTTNQCFSSVSVELPVV 450  
DB 465 QPELLGMRIGAVWVTGIILATFTNLDNSPKGAQERLGFAPFAMSTTFTYCAEAIPVFLQ 524  
QY 451 EKKLFTHEIYSGYVSVYFLGKLLDLPMTLPSIITFCIVYFMLGLKPKADAFVMM 510  
DB 525 EYIFMREYATNAYRRSSVLSQSIIS-IPALIVLSASFAATTFWAGLDGANGFFPPY 583  
QY 511 FTLMVMYASASSVALAIAAGQSVSVATLLMTICFVFMIFSGLLVNTTIASMLSMLOQ 570  
DB 584 FTILASFAGSSVFTFLSGVIPNVMLGFTVVVAILAYFLFSGFFISDRIPVYWLPHY 643  
QY 571 FSPRPGFTALQHNELGON---TCPGLNATGNP-----C 603  
DB 644 ISLVKPYEGVQLQNEF-QNPTRCFARGVQLFDNSPLGFEFFNDVKVNLKMSGVLGNNV 701  
QY 604 NYATC--TGEYLVKQGI-DLSPWG-LMKNHVALACMIVIFLTIA---YLKLLF 650  
DB 702 TAETCVTTGIDILKQGGITDISKNCLW-----ITVANGFFERVL 742

## RESULT 14

T02567  
N:Alternate names: protein F12L6.1  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 02-Mar-2001  
C:Accession: T02567; T00545; C84816  
R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,  
submitted to the EMBL Data Library, August 1998  
A:Description: Arabidopsis thaliana chromosome II BAC T16B24 genomic sequence.  
A:Reference number: Z14679  
A:Accession: T02567  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-740 <ROU>  
A:Cross-references: EMBL:AC004697; NID:G3402671; PIDN:AAC28975.1; PID:G3402672  
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kau  
submitted to the EMBL Data Library, July 1998  
A:Description: Arabidopsis thaliana chromosome II BAC F12L6 genomic sequence.  
A:Reference number: Z14168  
A:Accession: T00545  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-362 <ROW>  
A:Cross-references: EMBL:AC004218; NID:G3355463; PIDN:AAC27826.1; PID:G3355464  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umavam, L.; Tallon, L.;  
Eisen, J.A.; Salzborg, S.L.; Fraser, C.M.; Venter,  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487; PMID:10617197  
A:Accession: C84816  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-740 <STO>  
A:Cross-references: GB:AE002093; NID:G3402672; PIDN:AAC28975.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g39350; T16B24.1; F12L6.1  
A:Map position: 2  
C:Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1; ATP-B  
C:Keywords: ATP  
F;110-310/Domain: ATP-binding cassette homology <ABC>

Query Match 21.4%; Score 717.5; DB 1; Length 740;  
Best Local Similarity 28.0%; Pred. No. 4.7e-43;  
Matches 193; Conservative 120; Mismatches 294; Indels 83; Gaps 16;

QY 26 NDLKAFTEGAVLSFHNICVYVKL-----SGFLPKR-----KPVEKEILSNING 69  
DB 56 NDGYMETVFPVLSFDNLTVNSVRPKLDFRNLPFRRTEDPEIAQTARPKTKTLNNISG 115  
QY 70 IMKFG-LNALIGTGGKSSLLDVLAAKDPGSLGSDVLNGAPRAN-FKNSGVYVVD 127  
DB 116 ETRDGEIMAVLGASGSGKSTLDALANRIAKSGIKGTVKLNGETLQSRMLKVISAYVMD 175  
QY 128 DVVMGTLTVRENLFQSAALRLATMTNHEKNERINRVIQELGDKVADSKVGTQPIRGVS 187  
DB 176 DLLFPMITVEETLWFAEFRLPSLPSKSKKURVQALDQLGIRNAKTIIGDEGHRGIS 235  
QY 188 GGERKRTSIGMELITDPSILFLDEPTGLDSSANAVLLLKMSKQGRITIFSIHQPRY 247  
DB 236 GGERRAVSIGIDIHDPILFLDEPTSGLDSTAFWVVKVLRKIAQSGSIVMSIHQPSH 295  
QY 248 SIFKLPDSTLLASGRMLFHPGAQALGVFPESAGHYCEAYNNPADFLDI-----NG 300  
DB 296 RVLGGLDLRLIFLSRGHTVYSGSPASLPFRFTFGSPIPENRTEFALDILRELEGSAG 355  
QY 301 DSTAVALNREEDFKATEIIPSKQDKPLIEKLAIEYVNSFYKETKAEHLQ---LSCGK 357  
DB 356 TRGLIENK---KWQEMKKQSNRQFPLTPP-SSPYNLTLKEAIAASISRGKLVSGGS 410

Qy 358 -----KKKITTVEKEISYTTTFCCHLRWVSKRSFKNLLGNPOASIAQIITVTVLGVIGA 411  
Db 411 VAHGGAATTNTTLLAVAFANPMWIEIKTLKRSWLNRSQPELPGIRIASVUITGFIAT 470  
Qy 412 IYGLKNDSTGIONRAGVLFLLTNOCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFL 471  
Db 471 VFVRLDNPSPGVQERGLGFAPAFAMSTWFTYTCADALPVFLQERYIFKRETAYNAYRSSYVL 530  
Qy 472 GKLLSDLLPMTLPSIIFTCIVTFMGLKPKADAFVMMFTLMMVAYSASSMALAIAAQ 531  
Db 531 SHAIVS-PPSLIFSLVAFATTTWAGLDGLTGLFYCLIIILASFWSSGFVFLSGVV 589  
Qy 532 SVVSATLMTICFVFMPIPSGLLVNLTITIASWLSWLOQFSPRYGFTALQHNFLGQNF 591  
Db 590 PSMVLGTTIIVAILAYFLFESGFINRIPDIWIFHYMSLVKPYEAVLQNEFSDATK 649  
Qy 592 C--PGLNATGNP-----CNVATC--TGEVYLKQG-IDLSP 623  
Db 650 CFVRGVQIFDNTPLGELPEVMKLLIGTVSKSLGVTSSTTCLTGTSDILRQGVQLSK 709  
Qy 624 WGLMKNHVALACMIVIFLTIA---YLKLF 650  
Db 710 WN-----CLFITVAFGFFRILF 727

RESULT 15  
E96742  
probable ABC transporter F17M19.11 [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001  
C;Accession: E96742  
R;Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A;Reference number: A86141; MUID:21016719; PMID:11130712  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-609 <STC>  
A;Cross-references: GB:AE005173; NID:g6978921; PIDN:AAF34313.1; GSPDB:GN00141  
C;Genetics:  
A;Gene: F17M19.11  
A;Map position: 1  
C;Superfamily: fruit fly white protein; ATP-binding cassette homology

Query Match 21.3%; Score 712.5; DB 2; Length 609;  
Best Local Similarity 31.0%; Pred No 8e-43;  
Matches 190; Conservative 125; Mismatches 256; Indels 41; Gaps 16;

Qy 56 RKPVEKEILNINIMKPG-LNALIGPTGGKSLDLVLAARKDPSGLSDVLINGAPRP 114  
Db 23 RSTEERTILSGVTGMISPGEFMAVLGPSSGSKTLNNAVAGRLHGSNLTKILINDGKIT 82  
Qy 115 ANFKCNSGYVQDDVVMGTLTVRENLQFSAALBLATMTWHEKNEINRINQELGLDKVA 174  
Db 83 KQTLKRTGFFVAQDQLLPHLTVRETLVFLVALLRPLSLTRDVKLRAESVISBELGTRCE 142  
Qy 175 DSKVGTQFIRGVGGERKRTSIGMELITDPSILFDEPTTGLDSTANAVLLLLKEMSK- 233  
Db 143 NTVVGNTFIRGISGGERKRVSIHAELLINPSLLVLDPTSGLDATAALRLVQTLAGLAHG 202  
Qy 234 QGRTIIFSIHQPRYSIFKLFDSLTLLASGRIMPHGPAQEAALGYFESAGYHCEAYNMPADF 293  
Db 203 KGKTVVTSIHQPSRSRVQMFDTVLLISEGKCLFVGKGRDAMAYFESVGFAPFMPNPAF 262  
Qy 294 FLDIING--DSTAVALNREEDFKATEIIEPSKQDKPLIEKLABI-----YVNSSFYKETKA 347

Db 263 LLDLANGVCQTDGVTEREKENVRQTLVTAYDTLLAPQVKTCTIEVSHFPQDNARFVKT--- 319  
Qy 348 ELHQLSGGEKKKITTVEKEISYTTTFCCHLRWVSKRSFKNLLGNPOASIAQIITVTVL 406  
Db 320 ---RVNGGGITTCTA-----TWFSQCLDLHRLKKERRHESF-----DLIRIFQVVAAS 365  
Qy 407 LVIGAIYFGLKNDSTGIONRAGVLFLLTN--QCFSSVSVAVELFVVEKKLFIHEYISGYR 465  
Db 366 ILCGLMMW--HSDYRDVHDLGLIFFTISFWGLPSPFNNAVFTTPOERAIPTRRASGMYT 423  
Qy 466 VSSYFELGKLLSDLLPMTLPSIIFTCIVTFMGLKPKADAFVMMFTLMMVAYSASSMAL 525  
Db 424 LSSYFMARVLGSLSMELVLPASFLT-FTYMMVYLPRGIVFPFLLTSLVLLLYVLASQGLGL 482  
Qy 526 AIAAGQSVSVATLLMTICFVFMPIPSGLLVNLTITIASWLSWLOQFSPRYGFT---TALQ 582  
Db 483 ALGAAIMDAKKASTIVITVMTLAFVLGTGYVN--KVPFGVMVMKVVSTTFYCYELLVAIQ 540  
Qy 593 HNEFLGQNFPC--GLNATGNPNPNYATCTGEEYLVKQIDLSPLWGLMKNHVALACMIVIF 640  
Db 541 YGS--GEEILRMGLCDSKKGASAAATSAGCRFVEEVI--GDVGMWTSVGLVFLMPFGY 596  
Qy 641 LTIAYLKILFLK 652  
Db 597 RVLAYLALRRIK 608

Search completed: March 23, 2004, 18:31:36  
Job time : 23 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

QM protein - protein search, using sw model  
Run on: March 23, 2004, 18:31:13 ; Search time 55 Seconds  
(without alignments)  
3083.923 Million cell updates/sec

Title: US-09-961-086-1  
Perfect score: 3352  
Sequence: 1 MSSNVVEFIPVSGQNTNGF.....MIVFIETIAYLKLFLKKYS 655

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1049977 seqs, 258955339 residues  
Total number of hits satisfying chosen parameters: 1049977

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : Published Applications AA.\*
- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
  - 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
  - 3: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
  - 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
  - 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
  - 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
  - 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
  - 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
  - 9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
  - 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
  - 11: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
  - 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
  - 13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*
  - 14: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*
  - 15: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*
  - 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
  - 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
  - 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3352	100.0	655	10	US-09-961-086-1 Sequence 1, Appl
2	3352	100.0	655	15	US-10-405-806-13 Sequence 13, Appl
3	3346	99.8	655	9	US-09-961-353-35 Sequence 35, Appl
4	3346	99.8	655	14	US-10-120-687-61 Sequence 61, Appl
5	3346	99.8	655	15	US-10-405-806-2 Sequence 2, Appl
6	3338	99.6	655	9	US-09-866-866A-27 Sequence 27, Appl
7	3331	99.4	655	9	US-09-866-866A-10 Sequence 10, Appl
8	3331	99.4	655	14	US-10-090-455-5 Sequence 5, Appl
9	3053.5	91.1	604	9	US-09-745-763-197 Sequence 197, Appl
10	2757	82.2	657	9	US-09-866-866A-14 Sequence 14, Appl
11	835.5	24.9	1049	15	US-10-369-493-1520 Sequence 1520, Ap
12	812	24.2	663	13	US-10-108-605-245 Sequence 245, App
13	805	24.0	1084	12	US-10-424-599-242078 Sequence 242078,
14	805	24.0	1101	12	US-10-425-114-63125 Sequence 63125, A
15:	795.5	23.7	676	15	US-10-369-493-3799 Sequence 3799, Ap

16	773.5	23.1	695	12	US-10-424-599-176182 Sequence 176182,
17	746.5	22.3	819	12	US-10-425-114-54421 Sequence 54421, A
18	711	21.2	646	14	US-10-090-455-13 Sequence 13, Appl
19	710.5	21.2	674	14	US-10-090-455-4 Sequence 4, Appl
20	710.5	21.2	638	13	US-10-429-160-10 Sequence 10, Appl
21	710	21.2	638	13	US-10-072-621-10 Sequence 9, Appl
22	710	21.2	646	13	US-10-072-621-9 Sequence 2, Appl
23	710	21.2	646	14	US-10-090-455-2 Sequence 8, Appl
24	708	21.1	627	14	US-10-090-455-4 Sequence 4, Appl
25	708	21.1	646	13	US-10-154-452-4 Sequence 175941,
26	707	21.1	725	12	US-10-424-599-175941 Sequence 8, Appl
27	700	20.9	646	13	US-10-154-452-8 Sequence 53846, A
28	699	20.9	656	12	US-10-425-114-53846 Sequence 64380, A
29	699	20.9	673	12	US-10-425-114-64380 Sequence 1, Appl
30	693.5	20.7	652	9	US-09-837-992-1 Sequence 2, Appl
31	693.5	20.7	652	10	US-09-989-981A-2 Sequence 2, Appl
32	693	20.7	646	14	US-10-079-087-2 Sequence 14, Appl
33	683.5	20.4	599	15	US-10-210-130-14 Sequence 3, Appl
34	682.5	20.4	651	9	US-09-837-992-3 Sequence 6, Appl
35	682.5	20.4	651	14	US-10-090-455-6 Sequence 5740, Ap
36	682.5	20.4	651	14	US-10-090-455-6 Sequence 2025, Ap
37	677	20.2	559	15	US-10-369-493-5740 Sequence 4, Appl
38	664	19.8	608	15	US-10-369-493-5748 Sequence 154459,
39	658.5	19.6	1095	15	US-10-369-493-2025 Sequence 8, Appl
40	657.5	19.6	672	10	US-09-989-981A-4 Sequence 7, Appl
41	644	19.2	623	12	US-10-424-599-154459 Sequence 5347, Ap
42	640.5	19.1	673	10	US-09-989-981A-8 Sequence 6184, Ap
43	639.5	19.1	673	14	US-10-090-455-7
44	639	19.1	658	15	US-10-369-493-5347
45	636.5	19.0	639	15	US-10-369-493-6184

ALIGNMENTS

RESULT 1

US-09-961-086-1  
; Sequence 1, Application US/09961086  
; Publication No. US20030036645A1  
; GENERAL INFORMATION:  
; APPLICANT: UNIVERSITY OF MARYLAND, BALTIMORE  
; APPLICANT: ROSS, Douglas D.  
; APPLICANT: DOYLE, L. Austin  
; APPLICANT: ABRUZZO, Lynne  
; TITLE OF INVENTION: BREAST CANCER RESISTANCE PROTEIN (BCRP) AND THE DNA  
; TITLE OF INVENTION: WHICH ENCODES IT  
; FILE REFERENCE: EP19376-019  
; CURRENT APPLICATION NUMBER: US/09/961.086  
; PRIOR FILING DATE: 2001-09-21  
; PRIOR APPLICATION NUMBER: US 60/073,763  
; PRIOR FILING DATE: 1998-02-05  
; PRIOR APPLICATION NUMBER: PCT/US99/02577  
; PRIOR FILING DATE: 1999-02-05  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: Patent in ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 655  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-961-086-1

Query Match 100.0%; Score 3352; DB 10; Length 655;  
Best Local Similarity 100.0%; Pred. No. 7e-300;  
Matches 655; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MSSNVVEFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICVRVCLKSGFLPCRKPV 60
Db	1	MSSNVVEFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICVRVCLKSGFLPCRKPV 60
QY	61	KEILSNINGIMKPGINAILGPTGGKSSLLDLVAARXDPGSLGSDVLINGAPRANPKCN 120
Db	61	KEILSNINGIMKPGINAILGPTGGKSSLLDLVAARXDPGSLGSDVLINGAPRANPKCN 120

```
QY 121 SGVVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIOELGDKVADSKVGT 180
Db 121 SGVVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIOELGDKVADSKVGT 180
QY 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSDSTANAVLLLLKRMKSKQGRITIF 240
Db 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSDSTANAVLLLLKRMKSKQGRITIF 240
QY 241 SIHQPRYSIFKLFDSITLLASGRLMFHGPAQALGYFESAGYHCEAYNNPADFFLDIING 300
Db 241 SIHQPRYSIFKLFDSITLLASGRLMFHGPAQALGYFESAGYHCEAYNNPADFFLDIING 300
QY 301 DSTAVALNREEDFKATEIIEPSKQDKPLEKLAIEIVNSSFYKETAELHQLSGGEKKK 360
Db 301 DSTAVALNREEDFKATEIIEPSKQDKPLEKLAIEIVNSSFYKETAELHQLSGGEKKK 360
QY 361 ITVFKEISYTTSCFCHOLRWVSKRSFNLLGNPQASIAQIIVTVLGLVIGAIYFGLKND 420
Db 361 ITVFKEISYTTSCFCHOLRWVSKRSFNLLGNPQASIAQIIVTVLGLVIGAIYFGLKND 420
QY 421 TGIQNRAGVLFLLTNQCFSSYSAVELFVVEKKLFIHEIYISGYRVSSYFGLKLLSDLLP 480
Db 421 TGIQNRAGVLFLLTNQCFSSYSAVELFVVEKKLFIHEIYISGYRVSSYFGLKLLSDLLP 480
QY 481 MTMLPSIITCTIYVFMGLGPKADAFVMMFTLMVAYSASSMALAIAAGQSVSVATLL 540
Db 481 MTMLPSIITCTIYVFMGLGPKADAFVMMFTLMVAYSASSMALAIAAGQSVSVATLL 540
QY 541 MTICFVFMIFSGLLVNLTTIASWLSWQYFSIPRYGFTALQHNFLGQFCPGLNATGN 600
Db 541 MTICFVFMIFSGLLVNLTTIASWLSWQYFSIPRYGFTALQHNFLGQFCPGLNATGN 600
QY 601 NPCNATCTGEEYLVKQGDIDSPWGLWKNHVALACMIVFLTIAYLKLFLKKYS 655
Db 601 NPCNATCTGEEYLVKQGDIDSPWGLWKNHVALACMIVFLTIAYLKLFLKKYS 655
```

## RESULT 2

```
US-10-405-806-13
; Sequence 13, Application US/10405806
; Publication No. US2003023362A1
; GENERAL INFORMATION:
; APPLICANT: KOMATANI, HIDEYA
; APPLICANT: KARA, YOSHIKAZU
; APPLICANT: KOTANI, HIDEHITO
; APPLICANT: NAKAGAWA, RINAKO
; TITLE OF INVENTION: DRUG RESISTANT GENE AND USE THEREOF
; FILE REFERENCE: 234985USOCONT
; CURRENT APPLICATION NUMBER: US/10/405,806
; CURRENT FILING DATE: 2003-04-03
; PRIOR APPLICATION NUMBER: PCT/JP01/08112
; PRIOR FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: JP2000-303441
; PRIOR FILING DATE: 2000-10-03
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 13
; TYPE: PRT
; LENGTH: 655
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ABC02 482Tmutant sequence
US-10-405-806-13
```

```
Query Match 100.0%; Score 3352; DB 15; Length 655;
Best Local Similarity 100.0%; Pred. No. 7e-300;
Matches 655; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSSSNVEVFIPVSGQNTNGPPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPVE 60
Db 1 MSSSNVEVFIPVSGQNTNGPPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPVE 60
QY 61 KEILSNINGIMKPLGNAIIGPTGGKSSLLDVLAAKDPGSLGSDVLINGAPRANFKCN 120
```

```
Db 61 KEILSNINGIMKPLGNAIIGPTGGKSSLLDVLAAKDPGSLGSDVLINGAPRANFKCN 120
QY 121 SGVVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIOELGDKVADSKVGT 180
Db 121 SGVVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIOELGDKVADSKVGT 180
QY 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSDSTANAVLLLLKRMKSKQGRITIF 240
Db 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSDSTANAVLLLLKRMKSKQGRITIF 240
QY 241 SIHQPRYSIFKLFDSITLLASGRLMFHGPAQALGYFESAGYHCEAYNNPADFFLDIING 300
Db 241 SIHQPRYSIFKLFDSITLLASGRLMFHGPAQALGYFESAGYHCEAYNNPADFFLDIING 300
QY 301 DSTAVALNREEDFKATEIIEPSKQDKPLEKLAIEIVNSSFYKETAELHQLSGGEKKK 360
Db 301 DSTAVALNREEDFKATEIIEPSKQDKPLEKLAIEIVNSSFYKETAELHQLSGGEKKK 360
QY 361 ITVFKEISYTTSCFCHOLRWVSKRSFNLLGNPQASIAQIIVTVLGLVIGAIYFGLKND 420
Db 361 ITVFKEISYTTSCFCHOLRWVSKRSFNLLGNPQASIAQIIVTVLGLVIGAIYFGLKND 420
QY 421 TGIQNRAGVLFLLTNQCFSSYSAVELFVVEKKLFIHEIYISGYRVSSYFGLKLLSDLLP 480
Db 421 TGIQNRAGVLFLLTNQCFSSYSAVELFVVEKKLFIHEIYISGYRVSSYFGLKLLSDLLP 480
QY 481 MTMLPSIITCTIYVFMGLGPKADAFVMMFTLMVAYSASSMALAIAAGQSVSVATLL 540
Db 481 MTMLPSIITCTIYVFMGLGPKADAFVMMFTLMVAYSASSMALAIAAGQSVSVATLL 540
QY 541 MTICFVFMIFSGLLVNLTTIASWLSWQYFSIPRYGFTALQHNFLGQFCPGLNATGN 600
Db 541 MTICFVFMIFSGLLVNLTTIASWLSWQYFSIPRYGFTALQHNFLGQFCPGLNATGN 600
QY 601 NPCNATCTGEEYLVKQGDIDSPWGLWKNHVALACMIVFLTIAYLKLFLKKYS 655
Db 601 NPCNATCTGEEYLVKQGDIDSPWGLWKNHVALACMIVFLTIAYLKLFLKKYS 655
```

## RESULT 3

```
US-09-981-353-35
; Sequence 35, Application US/09981353
; Patent No. US20020160382A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy W.
; APPLICANT: Jones, David A.
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
; FILE REFERENCE: PA-0038 US
; CURRENT APPLICATION NUMBER: US/09/981,353
; CURRENT FILING DATE: 2001-10-11
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 35
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020160382A1 5517972CD1
US-09-981-353-35
```

```
Query Match 99.8%; Score 3346; DB 9; Length 655;
Best Local Similarity 99.8%; Pred. No. 2.5e-299;
Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MSSSNVEVFIPVSGQNTNGPPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPVE 60
Db 1 MSSSNVEVFIPVSGQNTNGPPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPVE 60
QY 61 KEILSNINGIMKPLGNAIIGPTGGKSSLLDVLAAKDPGSLGSDVLINGAPRANFKCN 120
Db 61 KEILSNINGIMKPLGNAIIGPTGGKSSLLDVLAAKDPGSLGSDVLINGAPRANFKCN 120
```

121 SGVVDVDMGTLTVRENLQFSAALRLATTMTNHEKNERINRVIOELGLDKVADSKVGT 180  
121 SGVVDVDMGTLTVRENLQFSAALRLATTMTNHEKNERINRVIOELGLDKVADSKVGT 180  
181 QFIRGSGGERKRTSIGMELITDPSILFDEPTTGLDSDSTANAVALLLKRMSKQGRITIF 240  
181 QFIRGSGGERKRTSIGMELITDPSILFDEPTTGLDSDSTANAVALLLKRMSKQGRITIF 240  
241 SHQPRYSIFKLFDSLTLLASGRLMFHPGPAQALGVFESAGVHCEAYNNPADFFLDIING 300  
241 SHQPRYSIFKLFDSLTLLASGRLMFHPGPAQALGVFESAGVHCEAYNNPADFFLDIING 300  
301 DSTAVALNREEDPKATEIIEPSKQDKPLIEKLAIEIVNSSFYKETAELHQLSGGKKKK 360  
301 DSTAVALNREEDPKATEIIEPSKQDKPLIEKLAIEIVNSSFYKETAELHQLSGGKKKK 360  
361 ITVFKEISYTTSFCHQLRWVSKRSFNKLLGNPOASIAQIIVTVLGLVIGAIYFGLKND 420  
361 ITVFKEISYTTSFCHQLRWVSKRSFNKLLGNPOASIAQIIVTVLGLVIGAIYFGLKND 420  
421 TGIONRAGVLFELTTNQCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFGLKLLSLLP 480  
421 TGIONRAGVLFELTTNQCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFGLKLLSLLP 480  
481 MTMLPSIIFTCIVYFVLMGLKPKADAFVVMFTLMWVAYSASSMALAIAAGQSVSVATLL 540  
481 MTMLPSIIFTCIVYFVLMGLKPKADAFVVMFTLMWVAYSASSMALAIAAGQSVSVATLL 540  
541 MTICFVFMIFSGLLVNLTTIASWSLQYFSIPRYGFTALQHNFLGQFCPLNATGN 600  
541 MTICFVFMIFSGLLVNLTTIASWSLQYFSIPRYGFTALQHNFLGQFCPLNATGN 600  
601 NPCNYATCTGEEYLVKQGDLSPWGLWKNHVALACMIVFLTIAYLKLLFLKKYS 655  
601 NPCNYATCTGEEYLVKQGDLSPWGLWKNHVALACMIVFLTIAYLKLLFLKKYS 655

RESULT 4  
US-10-120-687-61  
; Sequence 61, Application US/10120687  
; Publication No. US20030082155A1  
; GENERAL INFORMATION:  
; APPLICANT: Massachusetts General Hospital  
; TITLE OF INVENTION: Stem Cells of the Islets of Langerhans and Their Use in Treating  
; FILE REFERENCE: 3284/1235B  
; CURRENT APPLICATION NUMBER: US/10/120,687  
; CURRENT FILING DATE: 2002-04-11  
; PRIOR APPLICATION NUMBER: US60/169082  
; PRIOR FILING DATE: 1999-12-06  
; PRIOR APPLICATION NUMBER: US 09/963,875  
; PRIOR FILING DATE: 2001-09-25  
; PRIOR APPLICATION NUMBER: US 60/215109  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: US 60/238890  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: US 09/731261  
; PRIOR FILING DATE: 2000-12-06  
; NUMBER OF SEQ ID NOS: 61  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 61  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-120-687-61

Query Match 99.8%; Score 3346; DB 14; Length 655;  
Best Local Similarity 99.8%; Pred. No. 2.5e-299;  
Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
1 MSSNVEVFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPVE 60

Db 1 MSSNVEVFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPVE 60  
QY 61 KEILSNINGIMKPLGNAIIGPTGGKSSLLDLVAARKDPSGLSGDVLINGAPRANFKCN 120  
Db 61 KEILSNINGIMKPLGNAIIGPTGGKSSLLDLVAARKDPSGLSGDVLINGAPRANFKCN 120  
QY 121 SGVVDVDMGTLTVRENLQFSAALRLATTMTNHEKNERINRVIOELGLDKVADSKVGT 180  
Db 121 SGVVDVDMGTLTVRENLQFSAALRLATTMTNHEKNERINRVIOELGLDKVADSKVGT 180  
QY 181 QFIRGSGGERKRTSIGMELITDPSILFDEPTTGLDSDSTANAVALLLKRMSKQGRITIF 240  
Db 181 QFIRGSGGERKRTSIGMELITDPSILFDEPTTGLDSDSTANAVALLLKRMSKQGRITIF 240  
QY 241 SHQPRYSIFKLFDSLTLLASGRLMFHPGPAQALGVFESAGVHCEAYNNPADFFLDIING 300  
Db 241 SHQPRYSIFKLFDSLTLLASGRLMFHPGPAQALGVFESAGVHCEAYNNPADFFLDIING 300  
QY 301 DSTAVALNREEDPKATEIIEPSKQDKPLIEKLAIEIVNSSFYKETAELHQLSGGKKKK 360  
Db 301 DSTAVALNREEDPKATEIIEPSKQDKPLIEKLAIEIVNSSFYKETAELHQLSGGKKKK 360  
QY 361 ITVFKEISYTTSFCHQLRWVSKRSFNKLLGNPOASIAQIIVTVLGLVIGAIYFGLKND 420  
Db 361 ITVFKEISYTTSFCHQLRWVSKRSFNKLLGNPOASIAQIIVTVLGLVIGAIYFGLKND 420  
QY 421 TGIONRAGVLFELTTNQCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFGLKLLSLLP 480  
Db 421 TGIONRAGVLFELTTNQCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFGLKLLSLLP 480  
QY 481 MTMLPSIIFTCIVYFVLMGLKPKADAFVVMFTLMWVAYSASSMALAIAAGQSVSVATLL 540  
Db 481 MTMLPSIIFTCIVYFVLMGLKPKADAFVVMFTLMWVAYSASSMALAIAAGQSVSVATLL 540  
QY 541 MTICFVFMIFSGLLVNLTTIASWSLQYFSIPRYGFTALQHNFLGQFCPLNATGN 600  
Db 541 MTICFVFMIFSGLLVNLTTIASWSLQYFSIPRYGFTALQHNFLGQFCPLNATGN 600  
QY 601 NPCNYATCTGEEYLVKQGDLSPWGLWKNHVALACMIVFLTIAYLKLLFLKKYS 655  
Db 601 NPCNYATCTGEEYLVKQGDLSPWGLWKNHVALACMIVFLTIAYLKLLFLKKYS 655

RESULT 5  
US-10-405-806-2  
; Sequence 2, Application US/10405806  
; Publication No. US20030232362A1  
; GENERAL INFORMATION:  
; APPLICANT: KOMATANI, HIDEYA  
; APPLICANT: HARA, YOSHIKAZU  
; APPLICANT: KOTANI, HIDEHITO  
; APPLICANT: NAKAGAWA, RINAKO  
; TITLE OF INVENTION: DRUG RESISTANT GENE AND USE THEREOF  
; FILE REFERENCE: 234985USOCNT  
; CURRENT APPLICATION NUMBER: US/10/405,806  
; CURRENT FILING DATE: 2003-04-03  
; PRIOR APPLICATION NUMBER: PCT/JP01/08112  
; PRIOR FILING DATE: 2001-09-18  
; PRIOR APPLICATION NUMBER: JP2000-303441  
; PRIOR FILING DATE: 2000-10-03  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: Patentin version 3.2  
; SEQ ID NO 2  
; LENGTH: 655  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-405-806-2

Query Match 99.8%; Score 3346; DB 15; Length 655;  
Best Local Similarity 99.8%; Pred. No. 2.5e-299;  
Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
1 MSSNVEVFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPVE 60

```
Db 1 MSSNVVEFIPVQNTNGFPATASNDLKAFTEGAVLSFHNI CYRVKLSGFLPCRKPE 60
61 KEILSNINGIMKPGNALILGPTGGKSSLLDVLARKDPSGLSGDVLINGAPRANFKCN 120
61 KEILSNINGIMKPGNALILGPTGGKSSLLDVLARKDPSGLSGDVLINGAPRANFKCN 120
121 SGYVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIQELGDKVADSKVGT 180
121 SGYVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIQELGDKVADSKVGT 180
181 QIRGVSGGERKRTSIGMELITDPSILFDEPTTGLDSDSTANAVALLKRMKSKGRTIIF 240
181 QIRGVSGGERKRTSIGMELITDPSILFDEPTTGLDSDSTANAVALLKRMKSKGRTIIF 240
241 SIHQPRYSIFKLFDSLTLLASGRLMFHGPAQALGYFESAGYHCEAYNNPADFFLDIING 300
241 SIHQPRYSIFKLFDSLTLLASGRLMFHGPAQALGYFESAGYHCEAYNNPADFFLDIING 300
301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEIYVNSSFYKETAELHQLSGEKKKK 360
301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEIYVNSSFYKETAELHQLSGEKKKK 360
361 ITVFKESITTSFCHQLRWVSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKXDS 420
361 ITVFKESITTSFCHQLRWVSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKXDS 420
421 TGIQNRAGVLPFLTTNQCPSSVSAYVELFVVEKKLFIHEYISGYRVSSYFGLKLSDLLP 480
421 TGIQNRAGVLPFLTTNQCPSSVSAYVELFVVEKKLFIHEYISGYRVSSYFGLKLSDLLP 480
481 MTMLPSIIFTCTIVYFVLMGLKPKADAFFVVMFTLMVAVSASSMALAIAAGOSVVSATLL 540
481 MTMLPSIIFTCTIVYFVLMGLKPKADAFFVVMFTLMVAVSASSMALAIAAGOSVVSATLL 540
541 MTICFVFMVIFSGLLVNLTTIASWSLQYFSIPRYGFTALQHNFLQNCFCPLNATGN 600
541 MTICFVFMVIFSGLLVNLTTIASWSLQYFSIPRYGFTALQHNFLQNCFCPLNATGN 600
601 NPCNVATCTGEEYLVKQGDLSFVGLWKNHVALACMIVIFLTIAVILKLLFLKKYS 655
601 NPCNVATCTGEEYLVKQGDLSFVGLWKNHVALACMIVIFLTIAVILKLLFLKKYS 655
```

```
RESULT 6
US-09-866-866A-27
; Sequence 27, Application US/09866866A
; Patent No. US20020102244A1
; GENERAL INFORMATION:
; APPLICANT: Sorrentino, Brian
; APPLICANT: Schuetz, John
; TITLE OF INVENTION: A Method of Identifying and/or Isolating Stem Cells
; FILE REFERENCE: 1340-1-021CIP2
; CURRENT APPLICATION NUMBER: US/09/866,866A
; CURRENT FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: 09/584,586
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: PCT/US99/11825
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: 60/086,988
; PRIOR FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 27
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-866-866A-27
```

Query Match 99.6%; Score 3338; DB 9; Length 655;  
Best Local Similarity 99.7%; Pred. No. 1.4e-298; Indels 0; Gaps 0;  
Matches 653; Conservative 0; Mismatches 2;

```
QY 1 MSSNVVEFIPVQNTNGFPATASNDLKAFTEGAVLSFHNI CYRVKLSGFLPCRKPE 60
Db 1 MSSNVVEFIPVQNTNGFPATASNDLKAFTEGAVLSFHNI CYRVKLSGFLPCRKPE 60
QY 61 KEILSNINGIMKPGNALILGPTGGKSSLLDVLARKDPSGLSGDVLINGAPRANFKCN 120
Db 61 KEILSNINGIMKPGNALILGPTGGKSSLLDVLARKDPSGLSGDVLINGAPRANFKCN 120
QY 121 SGYVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIQELGDKVADSKVGT 180
Db 121 SGYVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIQELGDKVADSKVGT 180
QY 181 QIRGVSGGERKRTSIGMELITDPSILFDEPTTGLDSDSTANAVALLKRMKSKGRTIIF 240
Db 181 QIRGVSGGERKRTSIGMELITDPSILFDEPTTGLDSDSTANAVALLKRMKSKGRTIIF 240
QY 241 SIHQPRYSIFKLFDSLTLLASGRLMFHGPAQALGYFESAGYHCEAYNNPADFFLDIING 300
Db 241 SIHQPRYSIFKLFDSLTLLASGRLMFHGPAQALGYFESAGYHCEAYNNPADFFLDIING 300
QY 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEIYVNSSFYKETAELHQLSGEKKKK 360
Db 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEIYVNSSFYKETAELHQLSGEKKKK 360
QY 361 ITVFKESITTSFCHQLRWVSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKXDS 420
Db 361 ITVFKESITTSFCHQLRWVSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKXDS 420
QY 421 TGIQNRAGVLPFLTTNQCPSSVSAYVELFVVEKKLFIHEYISGYRVSSYFGLKLSDLLP 480
Db 421 TGIQNRAGVLPFLTTNQCPSSVSAYVELFVVEKKLFIHEYISGYRVSSYFGLKLSDLLP 480
QY 481 MTMLPSIIFTCTIVYFVLMGLKPKADAFFVVMFTLMVAVSASSMALAIAAGOSVVSATLL 540
Db 481 MTMLPSIIFTCTIVYFVLMGLKPKADAFFVVMFTLMVAVSASSMALAIAAGOSVVSATLL 540
QY 541 MTICFVFMVIFSGLLVNLTTIASWSLQYFSIPRYGFTALQHNFLQNCFCPLNATGN 600
Db 541 MTICFVFMVIFSGLLVNLTTIASWSLQYFSIPRYGFTALQHNFLQNCFCPLNATGN 600
QY 601 NPCNVATCTGEEYLVKQGDLSFVGLWKNHVALACMIVIFLTIAVILKLLFLKKYS 655
Db 601 NPCNVATCTGEEYLVKQGDLSFVGLWKNHVALACMIVIFLTIAVILKLLFLKKYS 655
```

```
RESULT 7
US-09-866-866A-10
; Sequence 10, Application US/09866866A
; Patent No. US20020102244A1
; GENERAL INFORMATION:
; APPLICANT: Sorrentino, Brian
; APPLICANT: Schuetz, John
; TITLE OF INVENTION: A Method of Identifying and/or Isolating Stem Cells
; FILE REFERENCE: 1340-1-021CIP2
; CURRENT APPLICATION NUMBER: US/09/866,866A
; CURRENT FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: 09/584,586
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: PCT/US99/11825
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: 60/086,988
; PRIOR FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 10
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-866-866A-10
```

Query Match 99.4%; Score 3331; DB 9; Length 655;  
Best Local Similarity 99.4%; Pred. No. 6.1e-298; Indels 0; Gaps 0;  
Matches 651; Conservative 1; Mismatches 3;

```

Qy 1 MSSNVEFIPVSOQNTNGFPATASNDLKAFTEGAVLSFHNIYRVKLSGFLPCRKEVE 60
Db 1 MSSNVEFIPVSOQNTNGFPATVNDLKAFTEGAVLSFHNIYRVKLSGFLPCRKEVE 60
Qy 61 KEILSNINGIMKPGNAILGPTGGKSSLLDLVAARKOPSGLSGVDVINGAPRANFKCN 120
Db 61 KEILSNINGIMKPGNAILGPTGGKSSLLDLVAARKOPSGLSGVDVINGAPRANFKCN 120
Qy 121 SGVVODDVMGTLTVRENLOPSAALRLATTNTNHEKNERINRVIOELGLDKVADSKVGT 180
Db 121 SGVVODDVMGTLTVRENLOPSAALRLATTNTNHEKNERINRVIOELGLDKVADSKVGT 180
Qy 181 QIRGVSGGERKRTSIGMELITDPSILFDEPTTGLDSTANAVLLLLKRMSKQRTIIF 240
Db 181 QIRGVSGGERKRTSIGMELITDPSILFDEPTTGLDSTANAVLLLLKRMSKQRTIIF 240
Qy 241 SIHQPRYSIFKLPDSLTLLASGRMLPHGPAQAEALGYFESAGYHCEAYNNPADFFLDIING 300
Db 241 SIHQPRYSIFKLPDSLTLLASGRMLPHGPAQAEALGYFESAGYHCEAYNNPADFFLDIING 300
Qy 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEIYVNSFYKETAELHQLSGGKKKK 360
Db 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEIYVNSFYKETAELHQLSGGKKKK 360
Qy 361 ITVFKISYTTSPCHOLRWYKSKSFKNLLGNPOASTAQIIVTVVLGLVIGAIYFGLKND 420
Db 361 ITVFKISYTTSPCHOLRWYKSKSFKNLLGNPOASTAQIIVTVVLGLVIGAIYFGLKND 420
Qy 421 TGIONRAGVLFLLTTCNOCFSSVSAVELFVVEKKLFIHEIYSGYRVSSYFGLKLLSLLP 480
Db 421 TGIONRAGVLFLLTTCNOCFSSVSAVELFVVEKKLFIHEIYSGYRVSSYFGLKLLSLLP 480
Qy 481 MTMLPSIIFTCIVYFMGLKPKADAFVVMFTLMMVAYSSMALAIAAGQSVSVATLL 540
Db 481 MTMLPSIIFTCIVYFMGLKPKADAFVVMFTLMMVAYSSMALAIAAGQSVSVATLL 540
Qy 541 MTICFVPMIIFSGLLVNLTTIASWLSWLOQFSIPRYGFTALQHNEFLGQNFCEGLNATGN 600
Db 541 MTICFVPMIIFSGLLVNLTTIASWLSWLOQFSIPRYGFTALQHNEFLGQNFCEGLNATGN 600
Qy 601 NPCNYATCTGEBYLVKQIDLSPLWGLMKNHVALACMIVIFLTTIAYLKLLFLKKYS 655
Db 601 NPCNYATCTGEBYLVKQIDLSPLWGLMKNHVALACMIVIFLTTIAYLKLLFLKKYS 655

RESULT 8
US-10-090-455-5
; Sequence 5, Application US/10090455
; Publication No. US20030027259A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Hongyun
; APPLICANT: Le Bihan, Stephane
; TITLE OF INVENTION: NOVEL ABCG4 TRANSPORTER AND USES THEREOF
; FILE REFERENCE: 100103.406
; CURRENT APPLICATION NUMBER: US/10/090.455
; CURRENT FILING DATE: 2002-03-01
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-090-455-5

Query Match 99.4%; Score 3331; DB 14; Length 655;
Best Local Similarity 99.4%; Pred. No. 6.1e-298;
Matches 651; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
Qy 1 MSSNVEFIPVSOQNTNGFPATASNDLKAFTEGAVLSFHNIYRVKLSGFLPCRKEVE 60
Db 1 MSSNVEFIPVSOQNTNGFPATVNDLKAFTEGAVLSFHNIYRVKLSGFLPCRKEVE 60

```

## RESULT 9

```

US-09-745-763-197
; Sequence 197, Application US/09745763
; Patent No. US20020065394A1
; GENERAL INFORMATION:

```

APPLICANT: Jacobs, Kenneth

McCoy, John M.

LaVallie, Edward R.

Collins-Racie, Lisa A.

Svans, Cheryl

Merberg, David

Treacy, Maurice

Spaulding, Vikki

TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES

ENCODING THEM

NUMBER OF SEQUENCES: 219

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: MA

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/745,763

FILING DATE: 18-Jun-2000

CLASSIFICATION: <Unknown>

## ATTORNEY/AGENT INFORMATION:

NAME: Sprunger, Suzanne A.  
REGISTRATION NUMBER: 41,323  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8284  
TELEFAX: (617) 876-5851

## INFORMATION FOR SEQ ID NO: 197:

SEQUENCE CHARACTERISTICS:  
LENGTH: 604 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 197:  
US-09-745-763-197

Query Match 91.1%; Score 3053.5; DB 9; Length 604;  
Best Local Similarity 91.9%; Pred. No. 2,1e-272;  
Matches 602; Conservative 0; Mismatches 2; Indels 51; Gaps 1;

QY 1 MSSNVVEFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPE 60  
DB 1 MSSNVVEFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPE 60

QY 61 KEILSNINGIMKPGNALIGPTGGKSSLLDVLAAKDPGSLGSDVLINGAPRANFKCN 120  
DB 61 KEILSNINGIMKPGNALIGPTGGKSSLLDVLAAKDPGSLGSDVLINGAPRANFKCN 120

QY 121 SGYVQDDVVMGTLVRENLPQSAALRLATTWNHEKNERINRVIOELGLDKVADSKVGT 180  
DB 121 SGYVQD-----VGT 129

QY 181 QTRGVSGGERKRTSGMELITDPSLFLDEPTTGLDSTANAVALLLRMSKQGRITIF 240  
DB 130 QTRGVSGGERKRTSGMELITDPSLFLDEPTTGLDSTANAVALLLRMSKQGRITIF 189

QY 241 SIHOPRYSIFKLDLSLTLASGLMFGHPAQALGYFESAGYHCEAYNNPADFFLDIING 300  
DB 190 SIHOPRYSIFKLDLSLTLASGLMFGHPAQALGYFESAGYHCEAYNNPADFFLDIING 249

QY 301 DSTAVALNREE-DFKATEIIEPSKQDKPLIEKLAETVYVNSSFYKETKAEHLQSGGEKKK 360  
DB 250 DSTAVALNREE-DFKATEIIEPSKQDKPLIEKLAETVYVNSSFYKETKAEHLQSGGEKKK 309

QY 361 ITVFEISYTTSFCHQLRWKRSFKNLLGNPQASIAQIIVTVLGLVIGAIYFGLKND 420  
DB 310 ITVFEISYTTSFCHQLRWKRSFKNLLGNPQASIAQIIVTVLGLVIGAIYFGLKND 369

QY 421 TGIONRAGVLPFLTTNQCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFGLKLSDLLP 480  
DB 370 TGIONRAGVLPFLTTNQCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFGLKLSDLLP 429

QY 481 MTMLPSIIFTCTIVYFVLMGLKPKADAFFVMMFTLMVAVSASSMALAIAGQSVSVATLL 540  
DB 430 MRMLPSIIFTCTIVYFVLMGLKPKADAFFVMMFTLMVAVSASSMALAIAGQSVSVATLL 489

QY 541 MTICFVFMIFSGLLVNLTTIASWSLQYFIPRYGTALOHNEFLQGNPCPLNATGN 600  
DB 490 MTICFVFMIFSGLLVNLTTIASWSLQYFIPRYGTALOHNEFLQGNPCPLNATGN 549

QY 601 NPCNATCTGEEYLVKQIDLSFWGLWKNHVALACMIVIFLTIAVLLKLLFLKKYS 655  
DB 550 NPCNATCTGEEYLVKQIDLSFWGLWKNHVALACMIVIFLTIAVLLKLLFLKKYS 604

## RESULT 10

US-09-866-866A-14  
Sequence 14, Application US/09866866A  
Patent No. US20020102244A1  
GENERAL INFORMATION:  
APPLICANT: Sorrentino, Brian  
APPLICANT: Schuetz, John  
TITLE OF INVENTION: A Method of Identifying and/or Isolating Stem Cells

FILE REFERENCE: 1340-1-021CIP2  
CURRENT APPLICATION NUMBER: US/09/866,866A  
CURRENT FILING DATE: 2001-08-30  
PRIOR APPLICATION NUMBER: 09/584,586  
PRIOR FILING DATE: 2000-05-31  
PRIOR APPLICATION NUMBER: PCT/US99/11825  
PRIOR FILING DATE: 1999-05-27  
PRIOR APPLICATION NUMBER: 60/086,988  
PRIOR FILING DATE: 1998-05-28  
NUMBER OF SEQ ID NOS: 27  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 14  
LENGTH: 657  
TYPE: PRT  
ORGANISM: Mus musculus  
US-09-866-866A-14

Query Match 82.2%; Score 2757; DB 9; Length 657;  
Best Local Similarity 81.5%; Pred. No. 5.5e-245;  
Matches 536; Conservative 51; Mismatches 67; Indels 4; Gaps 3;

QY 1 MSSNVVEFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPE 60  
DB 1 MSSNDHVLVPMQSRNNGLPRMNSRAVRTLAEGDVLSPHHITYRVKVSGL-VRKTYE 59

QY 61 KEILSNINGIMKPGNALIGPTGGKSSLLDVLAAKDPGSLGSDVLINGAPRANFKCN 120  
DB 60 KEILSNINGIMKPGNALIGPTGGKSSLLDVLAAKDPGSLGSDVLINGAPRANFKCN 119

QY 121 SGYVQDDVVMGTLVRENLPQSAALRLATTWNHEKNERINRVIOELGLDKVADSKVGT 180  
DB 120 SGYVQDDVVMGTLVRENLPQSAALRLATTWNHEKNERINTIIEKLEKLVADSKVGT 179

QY 181 QTRGVSGGERKRTSGMELITDPSLFLDEPTTGLDSTANAVALLLRMSKQGRITIF 240  
DB 180 QTRGVSGGERKRTSGMELITDPSLFLDEPTTGLDSTANAVALLLRMSKQGRITIF 239

QY 241 SIHOPRYSIFKLDLSLTLASGLMFGHPAQALGYFESAGYHCEAYNNPADFFLDIING 300  
DB 240 SIHOPRYSIFKLDLSLTLASGLMFGHPAQALGYFESAGYHCEAYNNPADFFLDIING 299

QY 301 DSTAVALNREE-DFKATEIIEPSKQDKPLIEKLAETVYVNSSFYKETKAEHLQSGGEKKK 359  
DB 300 DSSAVMLNREEQNEANKTEEPSKGEKPVIEUSEFYINSIAYGETKAEHLQSGGEKKK 359

QY 360 KITVFEISYTTSFCHQLRWKRSFKNLLGNPQASIAQIIVTVLGLVIGAIYFGLKND 419  
DB 360 GTSAFKEPVVTSFCHQLRWIARRSFKNLLGNPQASVAQLIVTVLGLIIGAIYFGLKYD 419

QY 420 TGIONRAGVLPFLTTNQCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFGLKLSDLL 479  
DB 420 AAGMQNRAGVLPFLTTNQCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFGLKNSDLL 479

QY 480 PMTMLPSIIFTCTIVYFVLMGLKPKADAFFVMMFTLMVAVSASSMALAIAGQSVSVATL 539  
DB 480 PMFELPSVIFTCTIVYFVLMGLKKTVDFAFFIMFTLINVAVTASSMALAIATQSVSVATL 539

QY 540 LMTICFVFMIFSGLLVNLTTIASWSLQYFIPRYGTALOHNEFLQGNPCPLNATGN 599  
DB 540 LMTIAFVFMIFSGLLVNLRTIGWLSWQYFIPRYGTALQYNEFLQGEPCPGNFVD 599

QY 600 NNPC--NYATCTGEEYLVKQIDLSFWGLWKNHVALACMIVIFLTIAVLLKLLFLKKYS 655  
DB 600 NSTCVNSYAICTGNEVLIHQIELSPWGLWKNHVALACMIIIFLTIAVLLKLLFLKKYS 657

## RESULT 11

US-10-369-493-1520  
Sequence 1520, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.





Db 605 EISCTSSN---TTCPSGKVIETNFSAADPLDYVGLAILVSRVLAAL 655

## RESULT 13

US-10-424-599-242078  
; Sequence 242078, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 242078  
; LENGTH: 1084  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_60623C.1.pap  
US-10-424-599-242078

Query Match 24.0%; Score 805; DB 12; Length 1084;  
Best Local Similarity 29.4%; Pred. No. 1.5e-64;  
Matches 217; Conservative 120; Mismatches 250; Indels 150; Gaps 17;

Qy 35 AVLSFHNICVRYVKLSGFLPCRKPVKEILSNINGIMKPG-LNALIGPTGGKSLLDVL 93  
Db 373 ASLAFENVSHLK-----GRQILSDISGAHVGELMAMGASGAGKTTFLDLL 420  
Qy 94 AARKDPSGLS-GDVLINGAP-RPANKCNKSGVVDVVMGTLTVRENLOFSAALRLATT 151  
Db 421 -ARKKRGTLGDFWVNGEKITDDDFSAIGFVDDQDNLPTLVTHETIMDSALLRLPKD 479  
Qy 152 MTNHEKNERINRVIQELGLDKVADSKVGTQ--FIRGSGGKRRKTSIGMELITDPSILFL 209  
Db 480 MSWAVKAQKVEDVEKQIGYHIRHQLIGSESGRGISGGERKRVGIACELVTSILFL 539  
Qy 210 DEPTTGLDSTA-NAVLLLLKMSKQGTIIIFSIHQPRYSIFKLPDSLTLLASGLMPHG 268  
Db 540 YBPTSLGDYAFNVEICLVTVKSNRTVWFTIHOQRSNIVALPDQLVLAQGRVTVSG 599  
Qy 269 PAQEAALGPESAGYHCEAYNNPADPFLDI-----INGDSTAVAL 307  
Db 600 PFSKQNYFDSIGYSCPPGFNADYLDVTMHASRPDNPDPEDSHTYGEDLDGRATVASS 659  
Qy 308 NREEDFKATEIIEPS-----KQDKPLIE----- 330  
Db 660 TMAVKSIPSIISNPSIEEGRPKNKRKSIROQOERQLFSRKSTDRAPASLRSDGSIL 719  
Qy 331 -----KLA-----EIVVNSFYKETAELH----- 350  
Db 720 DKDNKWKVQARQAQSSINEDGHDLPDPAGNNTNLDVLVSSYEASDVKAELHSDIESAI 779  
Qy 351 -----QLSGEKKKKITVFKESITYSFCHQLRWKSRKSPKLLGNPQASIAQI 399  
Db 780 SDASESNGNPLHAPNGNSGKRGYKPKMLPTQFL-----ILSRTRWNLRYNPQMLTHY 835  
Qy 400 IVTVVLGLVIGATYFGLKNDSTGIONRAGVLFLLTNNOCFSSVSAVELFVVEKLPFIHEY 459  
Db 836 ATAILAVPLGLFYGLTDLKGFQNLGQFFFLALFGFSLTSLTTFAPERLLFVREER 895  
Qy 460 ISQYRVSSYFLGLKSLDLPMTLPSIIFTCIVTFMLGLKPKADAPFVMMFTLMMVAYS 519  
Db 896 AKGYITFLAYLAKVDFDVPLRIPLPAIMGIVTFMTGLVPTAPEFGFKMLLLVFNLA 955  
Qy 520 ASSMALIAAGQSWSVATLLMTICFVFMWIFSGLLVNLTTIASLWLOFSPRYGFT 579  
Db 956 AMMICLFLGIVVRNQVANLLGVLMFLSLFGGLFHNETIPKGLLWLOSLSIHFGE 1015

Qy 580 ALQHNE-----FLGQNFCEPLNATGNPCNYATCTGEEYLVKQIGIDLSWGLWKNHVALA 634  
Db 1016 GLIVNEVRVYLSLVDDKY--GLN-----LEVFGSAILSSFGFDV--LALWRDVAGLG 1062  
Qy 635 CMIVIFLTITAYIKLLPL 651  
Db 1063 VVSGAFFVLGFAALHFL 1079

## RESULT 14

US-10-425-114-63125  
; Sequence 63125, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 63125  
; LENGTH: 1101  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: JC-CMLE01810066F10\_FLI.pap  
US-10-425-114-63125

Query Match 24.0%; Score 805; DB 12; Length 1101;

Best Local Similarity 29.4%; Pred. No. 1.5e-64;  
Matches 217; Conservative 120; Mismatches 250; Indels 150; Gaps 17;

Qy 35 AVLSFHNICVRYVKLSGFLPCRKPVKEILSNINGIMKPG-LNALIGPTGGKSLLDVL 93  
Db 390 ASLAFENVSHLK-----GRQILSDISGAHVGELMAMGASGAGKTTFLDLL 437  
Qy 94 AARKDPSGLS-GDVLINGAP-RPANKCNKSGVVDVVMGTLTVRENLOFSAALRLATT 151  
Db 438 -ARKKRGTLGDFWVNGEKITDDDFSAIGFVDDQDNLPTLVTHETIMDSALLRLPKD 496  
Qy 152 MTNHEKNERINRVIQELGLDKVADSKVGTQ--FIRGSGGKRRKTSIGMELITDPSILFL 209  
Db 497 MSWAVKAQKVEDVEKQIGYHIRHQLIGSESGRGISGGERKRVGIACELVTSILFL 556  
Qy 210 DEPTTGLDSTA-NAVLLLLKMSKQGTIIIFSIHQPRYSIFKLPDSLTLLASGLMPHG 268  
Db 557 YBPTSLGDYAFNVEICLVTVKSNRTVWFTIHOQRSNIVALPDQLVLAQGRVTVSG 616  
Qy 269 PAQEAALGPESAGYHCEAYNNPADPFLDI-----INGDSTAVAL 307  
Db 617 PFSKQNYFDSIGYSCPPGFNADYLDVTMHASRPDNPDPEDSHTYGEDLDGRATVASS 676  
Qy 308 NREEDFKATEIIEPS-----KQDKPLIE----- 330  
Db 677 TMAVKSIPSIISNPSIEEGRPKNKRKSIROQOERQLFSRKSTDRAPASLRSDGSIL 736  
Qy 331 -----KLA-----EIVVNSFYKETAELH----- 350  
Db 737 DKDNKWKVQARQAQSSINEDGHDLPDPAGNNTNLDVLVSSYEASDVKAELHSDIESAI 796  
Qy 351 -----QLSGEKKKKITVFKESITYSFCHQLRWKSRKSPKLLGNPQASIAQI 399  
Db 797 SDASESNGNPLHAPNGNSGKRGYKPKMLPTQFL-----ILSRTRWNLRYNPQMLTHY 852  
Qy 400 IVTVVLGLVIGATYFGLKNDSTGIONRAGVLFLLTNNOCFSSVSAVELFVVEKLPFIHEY 459

Db 853 AIAILLAVFLGFLYGLTDDKGFQNLGFFFFLLALFGFSTLSTTFAPERLLFVRER 912  
QY 460 ISGYVRSSYFLGLSLDLPMTLPSIITFCIVVFMLGLKPKADAFVMMETLMMVAYS 519  
Db 913 AKGYTPLAYYLAUVDFIVPURIUPPALMGICIVPMTGLVETAPEFGKFMLLLVFNLA 972  
QY 520 ASSMALIAAGOSVSVATLLMTICFVFMIFSGLLVNLTTIASLWSLQYFSPRYGFT 579  
Db 973 AAMICFLGIVVRNQGAVNLGLVLMVLSLLFGGLLNHETIPKGLLMQSLSIHPGFE 1032  
QY 580 ALOHNE-----FLQNFPCINATGNPCNYATCTGEEVLYKQGLDLPWGLKKNHVALA 634  
Db 1033 GLIVNEVRVLSLVDDKY--GLN-----IEVPGSAILSSFGFDV--LALWRDVAGLG 1079  
QY 635 CMIVIELTIAYLKLEFL 651  
Db 1080 VVSGAFVFLGFAALHFL 1096

RESULT 15  
US-10-369-493-3799  
; Sequence 3799, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 3799  
; LENGTH: 676  
; TYPE: PRT  
; ORGANISM: Neurospora crassa  
US-10-369-493-3799

Query Match 23.7%; Score 795.5; DB 15; Length 676;  
Best Local Similarity 31.2%; Pred. No. 5.4e-64;  
Matches 199; Conservative 107; Mismatches 218; Indels 113; Gaps 11;  
QY 61 KEILSNTINGMKPG-LNAILGPTGGKSLDLVLAARKDPSGLSGDVLNGAP-RPANFK 118  
Db 1 KEILSGIQGMAHGEVTAINGASGAGXTTFDLILARKNRKGQVSGDFYNGEKVSDPEYK 60  
QY 119 CNSGYVVQDDVVMGTLTVRENLOFSAAALRLATTMTNHEKNERINRVIOELGLDKVADSKV 178  
Db 61 NAVGFVQDETMPLTLVHETILNSALLRLPKDMTRAKEQVIEVEKQLGIYHIRDSLI 120  
QY 179 GTQ--FIRGVSGGERKETSIGMELITDPSILFDEPTTGLDSDTA-NAVLLLLKRMKOG 235  
Db 121 GSEEGKRGISGGKRRVGIACELVTSIFLDEPTSGLDAYNAYNVECLVTLAKTYK 180  
QY 236 RTTIFSHQPRYSTFLKFLDSITLLASGLRMLFHGPAQCALGYFESAGVHCBAYNPNADFFL 295  
Db 181 RTVIFTIHPQNSNIVALFDRLLILAAQKTVISGPHQCQYFDQIGYTCPPGFNIADYLV 240  
QY 296 DI-----INGDSTAVALNREDFKA-----TEIIEPS----- 322  
Db 241 DLTWHAGSTSSYDDGTLSDVGVSGPSSTRAVKSIASVSGVSGDSDSLVSSSSSRPNKR 300  
QY 323 -----KODKPL----- 328  
Db 301 RDSVRRQERELYTRRQAVDVTASSDAGDEIGYKLOKQPPVTFPLNSTNDDLHDLPLA 360  
QY 329 -----IEKLAEIYVNSSFYKETKAELHQL-----SGEKKKKITVFKETSYT----- 370

Db 361 ATGTDLDVLIESYIHSIDIAASTHEEIHQATAAAVNSGQNSGYVADGNI-YTGTMGKY 419  
QY 371 --TSFCHQLRWVSKRSGKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKNDSTGIONRAG 428  
Db 420 ARVGLFRQFVILSQRIWKNLRYRNPMLMTHYATAILLAVFAGLYFYGLTLDIAGFQNLG 479  
QY 429 VLPFLTTNQCFSSVSAVELFVVEKKLFIHEYISGYRVSVSYFLGKLLSDLLPMTMLPSII 488  
Db 480 LFFEVLLALFGFSTLTSGLVFSQBRLLFVRERANGYSPITYFAAKVLFDIVPLRIIPIL 539  
QY 489 FTCIVVFMGLKPKADAFVMMETLMMVAYSASSMALIAAAGOSVSVATLLMTICFVEM 548  
Db 540 LGALIYPMTGLVADYQRFVFLVLFVNLAAAACILFCILCKDGGVANLIGSLVWLFS 599  
QY 549 MIFSGLLVNLTTIASLWSLQYFSPRYGFTALOHNE 585  
Db 600 LLFAGLLLNHNAIPAAALWLQWLSIFHYGFEALIVNE 636

Search completed: March 23, 2004, 18:37:18  
Job time : 59 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2004, 18:28:24 ; Search time 23 Seconds  
(without alignments)  
1470.218 Million cell updates/sec

Title: US-09-961-086-1  
Perfect score: 3352  
Sequence: 1 MSSNVVEFIPVSGQNTNGF.....MIVIFLTAYLKLFLKKYS 655

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/2/iaa/5A-COMB.pep:\*  
2: /cgn2\_6/prodata/2/iaa/5B-COMB.pep:\*  
3: /cgn2\_6/prodata/2/iaa/6A-COMB.pep:\*  
4: /cgn2\_6/prodata/2/iaa/6B-COMB.pep:\*  
5: /cgn2\_6/prodata/2/iaa/PCITUS-COMB.pep:\*  
6: /cgn2\_6/prodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3352	100.0	655	US-09-245-808-1	Sequence 1, Appli
2	3331	99.4	655	US-09-767-594-1	Sequence 1, Appli
3	567	16.9	617	US-09-614-912-138	Sequence 138, App
4	537	16.0	1296	US-09-614-912-140	Sequence 140, App
5	461	13.8	539	US-09-614-912-144	Sequence 144, App
6	254	7.6	138	US-10-162-012-36	Sequence 36, Appl
7	246	7.3	270	US-09-134-001C-5561	Sequence 5561, Ap
8	240.5	7.2	653	US-09-543-681A-5411	Sequence 5411, Ap
9	238.5	7.1	1684	US-08-665-259-25	Sequence 25, Appl
10	238.5	7.1	1684	US-08-762-500-25	Sequence 25, Appl
11	238.5	7.1	1704	US-08-762-500-75	Sequence 75, Appl
12	232	6.9	256	US-09-134-001C-4600	Sequence 4600, Ap
13	228	6.8	321	US-09-107-532A-6490	Sequence 6490, Ap
14	227.5	6.8	273	US-09-134-001C-3841	Sequence 3841, Ap
15	226.5	6.8	1302	US-08-232-537-2	Sequence 2, Appli
16	224.5	6.7	382	US-09-134-001C-3439	Sequence 3439, Ap
17	222.5	6.6	322	US-09-107-532A-4662	Sequence 4662, Ap
18	222.5	6.6	345	US-09-107-532A-3849	Sequence 3849, Ap
19	222	6.6	1349	US-08-612-734B-2	Sequence 2, Appli
20	220	6.6	229	US-09-134-000C-3584	Sequence 3584, Ap
21	219	6.5	323	US-09-107-532A-4844	Sequence 4844, Ap
22	219	6.5	1457	US-08-665-259-27	Sequence 27, Appl
23	219	6.5	1457	US-08-762-500-27	Sequence 27, Appl
24	218.5	6.5	374	US-09-489-039A-8496	Sequence 8496, Ap
25	217.5	6.5	255	US-09-543-681A-6727	Sequence 6727, Ap
26	216.5	6.5	715	US-09-543-681A-7603	Sequence 7603, Ap
27	215	6.4	240	US-09-489-039A-12478	Sequence 12478, A

ALIGNMENTS

RESULT 1  
US-09-245-808-1  
; Sequence 1, Application US/9245808  
; Patent NO. 6313277  
; GENERAL INFORMATION:  
; APPLICANT: Doyle, L. Austin  
; APPLICANT: Abruzzo, Lynne V.  
; APPLICANT: Ross, Douglas D.  
; TITLE OF INVENTION: Breast Cancer Resistance Protein (BCRP) and DNA which  
; FILE REFERENCE: encodes it  
; FILE REFERENCE: Ross UMB conversion  
; CURRENT APPLICATION NUMBER: US/09/245,808  
; CURRENT FILING DATE: 1999-02-05  
; EARLIER APPLICATION NUMBER: 60/073763  
; EARLIER FILING DATE: 1998-02-05  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 655  
; TYPE: PRT  
; ORGANISM: Human MCF-7/AdrVp cells  
US-09-245-808-1

Query Match	100.0%;	Score	3352;	DB	4;	Length	655;
Best Local Similarity	100.0%;	Pred. NO.	0;				
Matches	655;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
Qy	1	MSSNVVEFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICVYVKLSGFLPCRKPYE	60				
Db	1	MSSNVVEFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICVYVKLSGFLPCRKPYE	60				
Qy	61	KEILSNINGIMKPGNAILGPTGGKSSLDVLAARNDKPSGLSGDVLINGAPRANPKCN	120				
Db	61	KEILSNINGIMKPGNAILGPTGGKSSLDVLAARNDKPSGLSGDVLINGAPRANPKCN	120				
Qy	121	SGYVQDDVVMGTLTVRENLQFSAALRLATNTNHEKNERINRVIQELGDKVADSKVGT	180				
Db	121	SGYVQDDVVMGTLTVRENLQFSAALRLATNTNHEKNERINRVIQELGDKVADSKVGT	180				
Qy	181	QFIRGVSGGGRKRTSIGMELITDPSILFLDEPTGLDSDSTANAVLLLLKRMKSQGRITIF	240				
Db	181	QFIRGVSGGGRKRTSIGMELITDPSILFLDEPTGLDSDSTANAVLLLLKRMKSQGRITIF	240				
Qy	241	SIHQPRYSIFKLPDLTLASGRIMFHPGPAQALGYEESAGYHCEAYNNPADFLDLING	300				
Db	241	SIHQPRYSIFKLPDLTLASGRIMFHPGPAQALGYEESAGYHCEAYNNPADFLDLING	300				
Qy	301	DSTAVLNREDDFKATEIIEPSKDKPLIEKLABIYVNSSFYKETKAEHLQLSGGEKKK	360				
Db	301	DSTAVLNREDDFKATEIIEPSKDKPLIEKLABIYVNSSFYKETKAEHLQLSGGEKKK	360				

QY 361 ITVFEISYTTSFCHQLRWVSKRSPKLLGNPQASIAQIIVTVVLGLVIGAIYFGLKND 420  
DB 361 ITVFEISYTTSFCHQLRWVSKRSPKLLGNPQASIAQIIVTVVLGLVIGAIYFGLKND 420  
QY 421 TGIQNRAGVLPFLTTNQCFSSVSAVELFVWEKLFHIEYISGYRVSSYFGLKLLDLP 480  
DB 421 TGIQNRAGVLPFLTTNQCFSSVSAVELFVWEKLFHIEYISGYRVSSYFGLKLLDLP 480  
QY 481 MTMLPSIIFTCTIVYFVPMGLKPKADAFFVMMFTLMMVAYSSASMAIAAGOSVSVATLL 540  
DB 481 MTMLPSIIFTCTIVYFVPMGLKPKADAFFVMMFTLMMVAYSSASMAIAAGOSVSVATLL 540  
QY 541 MTICFVFMIFSGLLVNTTITIASWLSWLOYSIPRYGFTALQHNFEFLGQFCPLGNATGN 600  
DB 541 MTICFVFMIFSGLLVNTTITIASWLSWLOYSIPRYGFTALQHNFEFLGQFCPLGNATGN 600  
QY 601 NPCYATCTGEBYLVKQIDISPMGLWKNHVALACMIVIFLTIAYLKLLFLKKYS 655  
DB 601 NPCYATCTGEBYLVKQIDISPMGLWKNHVALACMIVIFLTIAYLKLLFLKKYS 655

RESULT 2  
US-09-767-594-1  
; Sequence 1, Application US/09767594  
; Patent No. 6521635  
; GENERAL INFORMATION:  
; APPLICANT: Bates, Susan  
; APPLICANT: Robey, Robert  
; APPLICANT: The Government of the United States of America  
; APPLICANT: as represented by the Secretary of the  
; APPLICANT: Department of Health and Human Services  
; TITLE OF INVENTION: Inhibition of MXR Transport by Acridine Derivatives  
; FILE REFERENCE: 015280-402100US  
; CURRENT APPLICATION NUMBER: US/09/767,594  
; CURRENT FILING DATE: 2001-01-22  
; PRIOR APPLICATION NUMBER: US 60/177,410  
; PRIOR FILING DATE: 2000-01-20  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 655  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: human mitoxanthrone resistance (MXR)/BRCP/ABCP  
US-09-767-594-1

Query Match 99.4%; Score 3331; DB 4; Length 655;  
Best Local Similarity 99.4%; Pred. No. 0;  
Matches 651; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSSNVVEFIPVQNTNGFPATASNDLKATFEGAVLSFHNICVYRVKLSGFLPCRKPE 60  
DB 1 MSSNVVEFIPVQNTNGFPATVNSDLKATFEGAVLSFHNICVYRVKLSGFLPCRKPE 60  
QY 61 KEILSNGINGIMKPGNALIGPTGGKSLLDVLAARDPSGLSGDVLINGAPRANFKCN 120  
DB 61 KEILSNGINGIMKPGNALIGPTGGKSLLDVLAARDPSGLSGDVLINGAPRANFKCN 120  
QY 121 SGYVVQDDVVMGTLTVRENLFSAALRLATTMTNHEKNERINRVIELGLDKVADSKVGT 180  
DB 121 SGYVVQDDVVMGTLTVRENLFSAALRLATTMTNHEKNERINRVIELGLDKVADSKVGT 180  
QY 181 QFIRVSGGERKRTSIGNELITDPSILFDEPTGLDSDSTANAVILLKXMSKQGRITIF 240  
DB 181 QFIRVSGGERKRTSIGNELITDPSILFDEPTGLDSDSTANAVILLKXMSKQGRITIF 240  
QY 241 SIHQPRYSIFKLFDSLTLLASGRLMFHGPAQEALGFESAGYHCEAYNNPADFFLDIING 300  
DB 241 SIHQPRYSIFKLFDSLTLLASGRLMFHGPAQEALGFESAGYHCEAYNNPADFFLDIING 300

QY 301 DSTAVALNREEDPKATEIIEPSKQKPLIEKLABIYVNSSFYKETKABHQLSGGEKKK 360  
DB 301 DSTAVALNREEDPKATEIIEPSKQKPLIEKLABIYVNSSFYKETKABHQLSGGEKKK 360  
QY 361 ITVFEISYTTSFCHQLRWVSKRSPKLLGNPQASIAQIIVTVVLGLVIGAIYFGLKND 420  
DB 361 ITVFEISYTTSFCHQLRWVSKRSPKLLGNPQASIAQIIVTVVLGLVIGAIYFGLKND 420  
QY 421 TGIQNRAGVLPFLTTNQCFSSVSAVELFVWEKLFHIEYISGYRVSSYFGLKLLDLP 480  
DB 421 TGIQNRAGVLPFLTTNQCFSSVSAVELFVWEKLFHIEYISGYRVSSYFGLKLLDLP 480  
QY 481 MTMLPSIIFTCTIVYFVPMGLKPKADAFFVMMFTLMMVAYSSASMAIAAGOSVSVATLL 540  
DB 481 MTMLPSIIFTCTIVYFVPMGLKPKADAFFVMMFTLMMVAYSSASMAIAAGOSVSVATLL 540  
QY 541 MTICFVFMIFSGLLVNTTITIASWLSWLOYSIPRYGFTALQHNFEFLGQFCPLGNATGN 600  
DB 541 MTICFVFMIFSGLLVNTTITIASWLSWLOYSIPRYGFTALQHNFEFLGQFCPLGNATGN 600  
QY 601 NPCYATCTGEBYLVKQIDISPMGLWKNHVALACMIVIFLTIAYLKLLFLKKYS 655  
DB 601 NPCYATCTGEBYLVKQIDISPMGLWKNHVALACMIVIFLTIAYLKLLFLKKYS 655

RESULT 3  
US-09-614-912-138  
; Sequence 138, Application US/09614912  
; Patent No. 6677502  
; GENERAL INFORMATION:  
; APPLICANT: Allen, Steve  
; APPLICANT: Rafalski, Antoni  
; APPLICANT: Orozco, Buddy  
; APPLICANT: Miao, Gou-Hau  
; APPLICANT: Farnodu, Omolayo O.  
; APPLICANT: Lee, Jian Ming  
; APPLICANT: Sakai, Hajime  
; APPLICANT: Weng, Zude  
; APPLICANT: Caiml, Perry G  
; APPLICANT: Anderson, Shawn  
; TITLE OF INVENTION: Plant Metabolism Genes  
; FILE REFERENCE: B81378 US NA  
; CURRENT APPLICATION NUMBER: US/09/614,912  
; CURRENT FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: 60/143,401  
; PRIOR FILING DATE: 1999-07-12  
; PRIOR APPLICATION NUMBER: 60/143,412  
; PRIOR FILING DATE: 1999-07-12  
; PRIOR APPLICATION NUMBER: 60/146,650  
; PRIOR FILING DATE: 1999-07-30  
; PRIOR APPLICATION NUMBER: 60/170,906  
; PRIOR FILING DATE: 1999-12-15  
; PRIOR APPLICATION NUMBER: 60/172,959  
; PRIOR FILING DATE: 1999-12-21  
; PRIOR APPLICATION NUMBER: 60/172,946  
; PRIOR FILING DATE: 1999-12-21  
; NUMBER OF SEQ ID NOS: 204  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 138  
; LENGTH: 617  
; TYPE: PRT  
; ORGANISM: Zea mays  
US-09-614-912-138

Query Match 16.9%; Score 567; DB 4; Length 617;  
Best Local Similarity 28.0%; Pred. No. 8.4e-50;  
Matches 169; Conservative 117; Mismatches 224; Indels 94; Gaps 22;

QY 37 LSFHNCVYRVKLSGFLPCRKPEK-----EILSNGINGIMKPG-LNALIGPTGG 85  
DB 11 MSFDNNVYVDV-----PAEMKHQGVQDRLQLLREVTVGSPRPGVLTALMGVSGAG 61  
QY 86 KSSLDDVLAARDPSGLSGDVLINGAPR-PANFKCN SGYVVQDDVVMGTLTVRENLFQSA 144

Db 62 KTLMDVLGRGTGGYIEGDIHAGYKQATFARISGYCEQNDIHSQVTVRESLIYA 121  
QY 145 ALRLATMTNHE-----KNERINRVIQELGDKVADSKVGTQFIRGVSGGGRKTSIGME 199  
Db 122 FLRLPGKIGDQBITDDIKQFVDEWVWELVONLADALVGLPGITGLSTEQRKLTIAVE 181  
QY 200 LITDPSILFDEPTGLSDSTANAVLLLLKMSKQGRITIIHSIHOPRYSIFKLPSLTL 259  
Db 182 LVANPSIIPMDPTGLSDARAAAIWRTVTRNTVDTGRTVCTIHOPIIDIFESFDELL 241  
QY 260 -ASGRIMFHGP-----AQEALGYFESA-----GYHCEAYNNPADFFLDIINGSTAVA 306  
Db 242 KRGQVIYSGKGRNSQKWEYFEALPGVPKIKDKY-----NPATWMLV-----SSVATE 292  
QY 307 LNREDFKATEIIEPSKQDKPLIEKLABIYVNSPYKTKAELHOLSGEKKKKTIVPEKE 366  
Db 293 VLKMDP-----AKYETSPLYKQKLVNQLSQPEFGTSDLYP-P 332  
QY 367 ISYTSFCHLRWWSKRSFKNLLGNPQASIAQIIVTVVLGVIGAIYFGL-----KNDSTGI 423  
Db 333 TYSQSTIQPKACLWKQWLTWRSPDYNLVRSYFTLLVALLGSIFWRTGNMEDATLL 392  
QY 424 QNRAGVLF-----FLTNQCFSSVSAVELFVVEKKLFIHEYISGYTVRSYFLGKLLSLL 479  
Db 393 GMVIGAMTYAVNFIGINN- -STQPV- -VSIERTVYRERAGMYSAMFYAIAQVVE- I 448  
QY 480 PMTMLPSIIFTCIVVFMGLGKPKADA- - -PFVMMFTLMVAYSASSMALAIAAGQSVVS 535  
Db 449 PVYVQVOTTYTLIVVAMSFQWTAKEFFWFISFESLYFTY- YGMVVAISPNH- -E 504  
QY 536 VATLMTICFVMMIFSGLLVNLTTIASWLSWLOQFSI- - -PRYGTALQHNFLQON 590  
Db 505 VASIFAAAFSLNLFSGFFIPRIPGW- -WITYVICPLAWTVYGLIVTYQYGLDLELI 562  
QY 591 FCPG 594  
Db 563 SVPG 566

## RESULT 4

US-09-614-912-140  
; Sequence 140, Application US/09614912  
; Patent No. 6677502  
; GENERAL INFORMATION:  
; APPLICANT: Allen, Steve  
; APPLICANT: Rafalski, Antoni  
; APPLICANT: Orozco, Buddy  
; APPLICANT: Miao, Gou-Hau  
; APPLICANT: Famodu, Omolayo O.  
; APPLICANT: Lee, Jian Ming  
; APPLICANT: Sakai, Hajime  
; APPLICANT: Weng, Zude  
; APPLICANT: Caimi, Perry G  
; APPLICANT: Anderson, Shawn  
; TITLE OF INVENTION: Plant Metabolism Genes  
; FILE REFERENCE: BB1378 US NA  
; CURRENT APPLICATION NUMBER: US/09/614,912  
; CURRENT FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: 60/143,401  
; PRIOR FILING DATE: 1999-07-12  
; PRIOR APPLICATION NUMBER: 60/143,412  
; PRIOR FILING DATE: 1999-07-12  
; PRIOR APPLICATION NUMBER: 60/146,650  
; PRIOR FILING DATE: 1999-07-30  
; PRIOR APPLICATION NUMBER: 60/170,906  
; PRIOR FILING DATE: 1999-12-15  
; PRIOR APPLICATION NUMBER: 60/172,959  
; PRIOR FILING DATE: 1999-12-21  
; PRIOR APPLICATION NUMBER: 60/172,946  
; PRIOR FILING DATE: 1999-12-21  
; NUMBER OF SEQ ID NOS: 204  
; SOFTWARE: Microsoft Office 97

; SEQ ID NO 140  
; LENGTH: 1296  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-09-614-912-140

Query Match 16.0%; Score 537; DB 4; Length 1296;  
Best Local Similarity 27.5%; Pred.No. 3.9e-46; Indels 82; Gaps 22;  
Matches 162; Conservative 115; Mismatches 230;

QY 37 LSFHNICYRYKL-----KGFILPCRPVKEKEILSNINGIMKPG-LNAIILPTGGKSSLL 90  
Db 690 MSFDDVNYVYDMPAEKMQQGVDDR-----LQLLRDVTGSGFRPAVLTAALGVSAGKTTLM 745  
QY 91 DVLAAKDPGLSGDVLINGAP- PANFKCNSGYVVODDVWMTLVNRLQFSAALRLA 149  
Db 746 DVLAGRATGGYIEGDMRISGYPKQNTFARISGYCEQNDIHSQVTVRESLIYSAFLRP 805  
QY 150 TMTNHE-----KNERINRVIQELGDKVADSKVGTQFIRGVSGGGRKTSIGMELITDP 204  
Db 806 EKIGDQBITDDIKIQFVDEWVWELVONLADALVGLPGITGLSTEQRKLTIAVELVANP 865  
QY 205 SILFDEPTGLSDSTANAVLLLLKMSKQGRITIIHSIHOPRYSIFKLPSLTL-ASGR 263  
Db 866 SIIFMDEPTGLSDARAAAIWRTVTRNTVDTGRTVCTIHOPIIDIFEFDELLLRKGGQ 925  
QY 264 LMFHGP-----AQEALGYFESA-----GYHCEAYNNPADFFLDIINGSTAVLNREE 311  
Db 926 VIYSGQGRNSQKWEYFEALPGVPKIKDKY-----NPATWMLV-----SSVAAEVLNM 976  
QY 312 DFKATEIIEPSKQDKPLIEKLABIYVNSPYKTKAELHOLSGEKKKKTIVPEKEISYTT 371  
Db 977 DF-----AEYKTSPLYKQKLVNQLSQPEFGTSDLYP-PTKYSQ 1016  
QY 372 SFCHLRWWSKRSFKNLLGNPQASIAQIIVTVVLGVIGAIYF- -GLK-NDSTGIONRAG 428  
Db 1017 STGQFACLWKQWLTWRSPDYNLVRSYFTLLVALLGTFIPWKIGTKGNANSLRMVIG 1076  
QY 429 VLP-----FLTNQCFSSVSAVELFVVEKKLFIHEYISGYTVRSYFLGKLLSOLLPMTL 484  
Db 1077 AMTAVNFIGINN- -ATQPIVSIERTVYRERAGMYSAMFYAIAQVVE-IPYVVF 1132  
QY 485 PSIIFTCIVVFMGLGKPKADA- - -PFVMMFTLMVAYSASSMALAIAAGQSVVSVATLL 540  
Db 1133 QTAYTLIVVAMSFQWTAKEFFWFISFESLYFTY- YGMVVAISPNH- - -EVAALF 1188  
QY 541 MTICFVMMIFSGLLVNLTTIASWLSWLOQFSI-PR- - -YGFALQHNE 585  
Db 1189 AAAPYSLEFLNLFSGFFIPRIPGW- -WYWLCLPLAWTVYGLIVTYQYGD 1236

## RESULT 5

US-09-614-912-144  
; Sequence 144, Application US/09614912  
; Patent No. 6677502  
; GENERAL INFORMATION:  
; APPLICANT: Allen, Steve  
; APPLICANT: Rafalski, Antoni  
; APPLICANT: Orozco, Buddy  
; APPLICANT: Miao, Gou-Hau  
; APPLICANT: Famodu, Omolayo O.  
; APPLICANT: Lee, Jian Ming  
; APPLICANT: Sakai, Hajime  
; APPLICANT: Weng, Zude  
; APPLICANT: Caimi, Perry G  
; APPLICANT: Anderson, Shawn  
; TITLE OF INVENTION: Plant Metabolism Genes  
; FILE REFERENCE: BB1378 US NA  
; CURRENT APPLICATION NUMBER: US/09/614,912  
; CURRENT FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: 60/143,401  
; PRIOR FILING DATE: 1999-07-12  
; PRIOR APPLICATION NUMBER: 60/143,412

```
; PRIOR FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: 60/146,650
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 60/170,906
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: 60/172,959
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/172,946
; PRIOR FILING DATE: 1999-12-21
; NUMBER OF SEQ ID NOS: 204
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 144
; LENGTH: 539
; TYPE: PRT
; ORGANISM: Triticum aestivum
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (272)...(273)
US-09-614-912-144

Query Match      13.8%; Score 461; DB 4; Length 539;
Best Local Similarity 25.0%; Pred. No. 7.5e-39;
Matches 148; Conservative 107; Mismatches 241; Indels 96; Gaps 20;

QY 102 LSGDVLINGAP-RPANFKCNSGVVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNER 160
Db 4 IEGEITVSGYPKQPTFARISGYCEQNDIHSPHTIYESLVSFSAWLKLPDAEVDERRRMF 63
QY 161 INRVIOELGLDKVADSKVGTQFIRGVGGERKRTSIGMELITDPSILFLDPTTGLDSS 220
Db 64 IEEIMDLVELSLRGALVCLPGVNLSTEQKRLTIAVELVANPSIIFMDPTSGLDARA 123
QY 221 ANAVLLKKRMSKQRTIIFSTHOPRYSIFKLFSLTL-ASGRLMFHPG-----AQBALG 275
Db 124 RAIVKRTVNTVNTGRTVCTIHQSDIFEFAPDLFLKKGGEIYVGPVGQNSANLIE 183
QY 276 YPESA-----GYHCEAYNPADFFLDIINGDSTAVALNREDFKATEIIEPSKQDKP 327
Db 184 YFEETEGISKIDGY-----NPATWLE-----VSSSAQDEM 215
QY 328 LIEKLAIEYVNSFFKYEKAEHLQSLGEGKKKTIYFKESYTSF---CHQLRWVKRS 384
Db 216 LGIDFAEVROSELYORNELKELKELMSPAGSSDLNF-PTQYSRSFVTCIACIWKQXXS 274
QY 385 FKNLLGNQASIAQIIVTVGLVIGAIYFGLKNDSTGIQN-----RAGVLFFLTIN 436
Db 275 YWR---NPSYTAIVLLFIIVIALMFGTWFMDLGSKTRESQDLFNAMGMYAAVLYIGVQN 331
QY 437 QCFSSVSVELFVVEKKLFIEHYISGYRVSSYFLGKLLSLLPMTLPSIIFTCIYFM 496
Db 332 ----SGSQVQVNVVVERTVYRERAAAGMYSAPFYAGQVAIE-PPYVLVQALYIGGLVYSM 386
QY 497 LGLKPKADAF----FVMGFTLMWVAYSASSMALAAGQSVSVVATLLMTICFVFMIFS 552
Db 387 IGFEVTVAKLWLYLFNFTMYLFF-YGMVAVGLTNE---STAALISSAFYVNWLF 442
QY 553 GLLV---NLTTIASLWMLQVPSIPRYGFTALQHNFEIQQNFCPLNATGNPNYATCT 609
Db 443 GYLIIPRKLPIWRRYSIMCPVANTLYGLVASQFGDI-----QQPLDQGV-P 488
QY 610 GBEYLVKQGI-----DLSPWGLKXKHHVALACMIVIFLAIYKLLFLAK 653
Db 489 GPQIIVAFVTDYFGFHHDFL-WVVMVHVAVFTVLFAELFSFAMRNFQKR 539

RESULT 6
US-10-162-012-36
; Sequence 36, Application US/10162012
; Patent No. 6682597
; GENERAL INFORMATION:
; APPLICANT: Curtis, Rory A.J.
; APPLICANT: Silos-Santiago, Inmaculada
; APPLICANT: Gu, Wei
```

RESULT 7  
 US-09-134-001C-5561  
 ; Sequence 5561, Application US/09134001C  
 ; Patent No. 6380370  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn Doucette-Stamm et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
 ; FILE REFERENCE: GTC-007  
 ; CURRENT APPLICATION NUMBER: US/09/134,001C  
 ; PRIOR FILING DATE: 1998-08-13  
 ; PRIOR APPLICATION NUMBER: US 60/064,964  
 ; PRIOR FILING DATE: 1997-11-08  
 ; PRIOR APPLICATION NUMBER: US 60/055,779  
 ; PRIOR FILING DATE: 1997-08-14  
 ; NUMBER OF SEQ ID NOS: 5674  
 ; SEQ ID NO 5561  
 ; LENGTH: 270  
 ; TYPE: PRT  
 ; ORGANISM: Staphylococcus epidermidis  
 US-09-134-001C-5561

Query Match 7.3%; Score 246; DB 4; Length 270;  
 Best Local Similarity 27.1%; Pred. No. 6e-17;  
 Matches 81; Conservative 56; Mismatches 102; Indels 60; Gaps 12;  
 QY 42 ICYRVKLV-----KSGFLPC-----RKPYEKEILSNINGIMKPG-LNALIGPTG 83  
 DB 1 VCLRKIIWLKYDRSDYLSLEVRQLSKYVGDKQAFQVLEINLDVDEGEFISMGPSG 60  
 QY 84 GKSSLLDVLARKDPGSLGSDVING--APRAN-----FKNSGVVQDDVVMGTLT 135  
 DB 61 SGKTLNLNLSIDYMT--KGSITINGKQLEKSNKQLSDIRKDKDIGIFQDYNILTLT 118  
 QY 136 VRENIQFSAALRLATMTNHEKNERINRVIOELGLDKVADSKVGTQFIRGVSGGERKRTS 195  
 DB 119 VKENIMPLSVQKLDQKINH--ERYQRIVEALNISDISD-----KYPSELGGGQRTS 170  
 QY 196 IGMEILTDPSILPLDEPTGLDSSTANAVLLLKRMKQKQ-GRTIIFSHQPRYSIFKLF 254  
 DB 171 AARAFINLPSIIFADEPTGALDSKSTLDLLKELKYNBEEFTILMLVTHDPVAASFS--N 228  
 QY 255 SLTLASGLMFGHQAQALGVFESAGYHCEAYNNPAD---PFLDIINGDSTAVALNRE 310  
 DB 229 RVNMLKGGQIP-----TELYQGGDDKQTFYKEIIRTSQVLGGINYE 269

RESULT 8  
 US-09-543-681A-5411  
 ; Sequence 5411, Application US/09543681A  
 ; Patent No. 6605709  
 ; GENERAL INFORMATION:  
 ; APPLICANT: GARY BRETON  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
 ; FILE REFERENCE: 2709.1002-001  
 ; CURRENT APPLICATION NUMBER: US/09/543,681A  
 ; CURRENT FILING DATE: 2000-04-05  
 ; PRIOR APPLICATION NUMBER: US 60/128,706  
 ; PRIOR FILING DATE: 1993-04-09  
 ; NUMBER OF SEQ ID NOS: 8344  
 ; SEQ ID NO 5411  
 ; LENGTH: 653  
 ; TYPE: PRT  
 ; ORGANISM: Proteus mirabilis  
 US-09-543-681A-5411

Query Match 7.2%; Score 240.5; DB 4; Length 653;  
 Best Local Similarity 20.1%; Pred. No. 1e-15;  
 Matches 130; Conservative 106; Mismatches 223; Indels 189; Gaps 27;

QY 55 CRKPEKEI--LSNINGIMKPG-----LNAILGPTGGKSSLDVLA 94  
 DB 2 CRRVMSALLENNVSRLYTNGEEDFVLDQLSLTNAGEMVAIIIGASGSGKSTLNNILG 61  
 QY 95 ARKDPGSLGSDVING-----APRANFKNSGVVQDDVVMGTLTVRENLF 142  
 DB 62 CLDKPS--SGEYKVGQCVADMESDQALALREHF---GFIFRYHLMAHLTAEQNVEI 115  
 QY 143 SAALRLATMTNHEKNERINRVIOELGLDKVADSKVGTQFIRGVSGGERKRTSIGMELIT 202  
 DB 116 PA---TYAGKSTEQRKERARALLTRGLAERTHYRPSQ-----LSGGQORVSIARALWN 167  
 QY 203 DPSILFDEPTGLDSSTANAVLLLKRMKQKRTIIFSHQPRYSIFKLFPSLTLLASG 262  
 DB 168 GGEVILADEPTGALDSQSGKEVMAILKQLNQOQHTVIVTHDP--LIAQQAIRIIEIKDG 225  
 QY 263 RLM-----PHG-----PAQALGVFESA-GVHCEAYNNP----- 290  
 DB 226 QIISDNNNHSAPVKVPPAIQTASYFHOVIGRFTQALNMAWRAMVNVKIRTLTLMGLI 285  
 QY 291 -----ADFFLDIINGDSTAVALNREDFKATEIIBPSKQDK-PLIEKLAIEYV 337  
 DB 286 IGIASVVTVIIVIGDAAKDRVLADIKAIGANTIDIYPKELGSDSPEDKQSLTIQDVDAK 345  
 QY 338 NSSFYKETKAEHL--QLSGEKKKKITV-----FKELIS-YTTSFCH---Q 376  
 DB 346 QQSYIOSVTPQIYFSRLRRGNQDAPATVSGVNEDYFVSVALKPAQGSTFTPDMIHRAQ 405  
 QY 377 LRWVSKRSFKNLGNPQASIA-QIIVTVVLGLVIGAI-----YFGLKND-----S 420  
 DB 406 VVVIDENTRHRFPFNQAVIGEQLIIRNIPSTIIGVVAEQKSTFGDNKSLRVWVPYSTLS 465  
 QY 421 TGTQNR-----GVLFELT-----TNOCFSSVSVELFVVE 451  
 DB 466 SRTYNSYLDNITVKYEGYDASVAPQQLRLTLIRHGKDIFTYNIDSFIRAAEKTOT 525  
 QY 452 KCLF-----IHEYISGYRVSSYFL-----GKLLSDLLPMTMLPSIIFT 490  
 DB 526 MQLFLTIVAVISLVGIGVGMNIMLVSTERTREIGIRMAVGARASDVMMQOPLIESVL-V 584  
 QY 491 CIVYFMGLKPKADAFVVMFTLMVY-----AYSASSMALATAAGOSV 533  
 DB 585 CLVGGLLGI---SLSPAIAFMASMLPNWHFVFPQTALISAFACSTAI 629

RESULT 9  
 US-08-665-259-25  
 ; Sequence 25, Application US/08665259  
 ; Patent No. 6028173  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Landes, Gregory M.  
 ; APPLICANT: Burn, Timothy C.  
 ; APPLICANT: Connors, Timothy D.  
 ; APPLICANT: Dackowski, William R.  
 ; APPLICANT: Van Raay, Terence J.  
 ; APPLICANT: Klingner, Katherine W.  
 ; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,  
 ; NUMBER OF SEQUENCES: 73  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: GENZYME CORPORATION  
 ; STREET: One Mountain Road  
 ; CITY: Framingham  
 ; STATE: Massachusetts  
 ; COUNTRY: United States of America  
 ; ZIP: 01701  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/665,259



FILING DATE: 17-JUN-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Dugan, Deborah A.  
REGISTRATION NUMBER: 37,315  
REFERENCE/DOCKET NUMBER: IGS-9.1  
TELEPHONE: (508) 872-8400  
TELEFAX: (508) 872-5415  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1684 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-665-259-25

Query Match 7.1%; Score 238.5; DB 3; Length 1684;  
Best Local Similarity 20.3%; Pred. No. 8e-15;  
Matches 169; Conservative 116; Mismatches 241; Indels 305; Gaps 39;

QY 66 NINGMKPLNALIGPTGGKSSLLDLVAARKDPSGLSGDVLING---APRPANFKNSG 122  
DB 533 NLN-LYEGQITVLLGHNGAGKTTLSMLTGLFPPT--SGRAYISGVEISQDMVQIRKSLG 569  
QY 123 YVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIOELGLDKVADSKVGTQF 182  
DB 590 LCPQHDIIFDNLTVAEHLFYFAQLK---GLSRQKCPPEVKQMLHIGLEKWNRSR----- 641  
QY 183 IRGVSGBGKRKTSIGMELITDPSILFDEPTTGLDSTANAVLLLLKRMKQGRITIFSI 242  
DB 642 SRFLSGGNRRKLSIGIALIAGSKVLLDEPTSGMDAISRAIWDLLQR-QKSDRTIVLT 700  
QY 243 HOPRYSIFK-----LFDLSLTLLASGRMLFHGPA---OEALGYFESAGVHC-----EAYNPN 290  
DB 701 H-----FWDERDLGLDRIAMKELQCCSSFLKQYK-----AGYHMLVKEPCNP 750  
QY 291 AD-----FFLD----- 296  
DB 751 EDISOLVHHVFNATLESSAGAEISFILPRESTHREGLFAKLEKKOKELGIASFASIT 810  
QY 297 -----IING 300  
DB 811 TWEVFLRVGLVDSDMDIQALPALQYQHERRASDWDVNSLGCAMDPSDGIKALIEE 870  
QY 301 DSTAVALN-----REDFKATEIIEP-----SKQ- 324  
DB 871 ERTAVKLTGLALHQQFWMFLKKAAYSWREKMAVAQVLVPLTCTVTLALLAINYSSEL 930  
QY 325 -DKPLI-----EKLA-----IYVNSSFYKFKAEHLQ----- 351  
DB 931 FDPMLRLTLGEYGTVPVFPVPGTSQIQQLSEHLKDALQAEQEPREVLGLDEEFLIF 990  
QY 352 ---LSGGBKKKITVFEKISYTSFCHOLRWVSKRSFKNLGNPQ-----ASIAQI 399  
DB 991 RASVEGGGFNERCLV-----AASF-----RDVGERTVVALFNQAYHSPATALAVVDNL 1040  
QY 400 IVTVVLGLVIGAIYGLKNDSTGIC-----NRAGVLFPLTNQCF-----SUSAV 445  
DB 1041 LFKLLCGPHASIVVSNFPQPSALQAQKQFNEGRKGFIDIALNLLFAMAFIATFSTILAV 1100  
QY 446 ELFVVEKELFIH-EYISGYRVYSSYFLGKLLSDLLPMTMLPSIIFTCIVYFMLGLKPKAD 504  
DB 1101 SERAVQAK---HVQFVSGVH-VASFWSALLWDLISF-LIFSLLL-LVVF-----X 1145  
QY 505 AFVVMFLMVAAYASASSWALIAAGQSVSVATLLMTICVFMFISGLLVNUTTIASW 564  
DB 1146 AFDVRAFT--RDGHVADTLLLLGVAIIPLMYLMN-----PFLGAATAYTALTIFNI 1198  
QY 565 LSWLOYF-----SIPRYGF-----TALQNEFLGONFCPGINATCNPNVAT---CTGE 611  
DB 1199 LSGIATFLMVTIMRIPAVKLEBSLKTLDHVLVLPNHCLGM-AVSSFVENYETRYCTSS 1257

QY 612 E-----YLVKQIDILSPW---GLWKNHVALACMIVIFLITAYLKLFL 651  
DB 1258 EVAHYCKKNIQOENFYAWSAPGVGRFVMAAS-----GCAYLILFL 1303

RESULT 10  
US-08-762-500-25  
Sequence 25, Application US/08762500  
Patent No. 6030806  
GENERAL INFORMATION:  
APPLICANT: Landes, Gregory M.  
APPLICANT: Burn, Timothy C.  
APPLICANT: Connors, Timothy D.  
APPLICANT: Dackowski, William R.  
APPLICANT: Van Raay, Terence J.  
APPLICANT: Klinger, Katherine W.  
TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,  
TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME  
NUMBER OF SEQUENCES: 83  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENZYME CORPORATION  
STREET: One Mountain Road  
CITY: Framingham  
STATE: Massachusetts  
COUNTRY: United States of America  
ZIP: 01701  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/762,500  
FILING DATE: 09-DEC-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/665,259  
FILING DATE: 17-JUN-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/10469  
FILING DATE: 17-JUN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Dugan, Deborah A.  
REGISTRATION NUMBER: 37,315  
REFERENCE/DOCKET NUMBER: IGS-9.3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (508) 872-8400  
TELEFAX: (508) 872-5415  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1684 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-762-500-25

Query Match 7.1%; Score 238.5; DB 3; Length 1684;  
Best Local Similarity 20.3%; Pred. No. 8e-15;  
Matches 169; Conservative 116; Mismatches 241; Indels 305; Gaps 39;

QY 66 NINGMKPLNALIGPTGGKSSLLDLVAARKDPSGLSGDVLING---APRPANFKNSG 122  
DB 533 NLN-LYEGQITVLLGHNGAGKTTLSMLTGLFPPT--SGRAYISGVEISQDMVQIRKSLG 569  
QY 123 YVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIOELGLDKVADSKVGTQF 182  
DB 590 LCPQHDIIFDNLTVAEHLFYFAQLK---GLSRQKCPPEVKQMLHIGLEKWNRSR----- 641  
QY 183 IRGVSGBGKRKTSIGMELITDPSILFDEPTTGLDSTANAVLLLLKRMKQGRITIFSI 242  
DB 642 SRFLSGGNRRKLSIGIALIAGSKVLLDEPTSGMDAISRAIWDLLQR-QKSDRTIVLT 700  
QY 243 HOPRYSIFK-----LFDLSLTLLASGRMLFHGPA---OEALGYFESAGVHC-----EAYNPN 290

Db 701 H-----FDEADLLGDRJAJMAKGLQCGSSFLKQYK-----AGYHMTLVKEPHCNP 750  
QY 291 AD-----FPLD-----296  
Db 751 EDISQLVHHVHPNATLESSSAGAELSFILPRESTRPEGLFAKLEKKQKELGSIASFGASIT 810  
QY 297 -----IING 300  
Db 811 TMEEVFLRVKGLVDSSMDIQAIQLPALQYQHERRASDAVDSNLCGAMPDPSDGIGALIEE 870  
QY 301 DSTAVALN-----REEDFKATEIIEP-----SKQ- 324  
Db 871 ERTAVKLTGLALHCOQFWAMFLKKAAYSWEKRWAAQVLVPLTCVTLLALAINYSSEL 930  
QY 325 -DKPLI-----EKLAEE-----IYVNSSFFKETAELHQ-----351  
Db 931 FDPMLRLTLGEYGRVTVVPSVPGTSQQLQSEHLKDALQAEQEPREVLGDLEEFLLIF 990  
QY 352 ---LSGGEKKKKTIVFKEISYTTSPCHQLRWKSKRSFKNLLGNPQ-----ASIAQI 399  
Db 991 RASVEGGGFNERCLV-----AASF-----RDVGERTVNNALFNNQAYHSPATALAVVDNL 1040  
QY 400 IVTVVLGLVIGAYFGLKNDSTGIQ-----NRAGVLFFLTNNQCFSS-----SVSAV 445  
Db 1041 LFKLLCGPHASIVVSNFPQPSRSALQAQKQFNEGRKGFIALNLLFAMAFLASTFSILAV 1100  
QY 446 ELFVVEKKKLFTH-EYISGYRVSSYFGLKLLSDLEMTMLPSIIFTCIVVFMGLKPKAD 504  
Db 1101 SERAVQAK---HVQFVSGVH-VASFWSALLNDLISF-LIPSLLL--LVVF-----X 1145  
QY 505 AFFVMMFTLMVAVSASSMALATAAGOSVVSATLMTICFVFMIFSGLLVNLTIASW 564  
Db 1146 AFDVRAFT--RDGHMADTLLLLLYGWAIPLMYLNN-----FFFLGATATVTRLTIFI 1198  
QY 565 LSWLQYF-----SIPRYGE-----TALOHEFLGQFCEGLNATGNPCNYAT---CTGE 611  
Db 1199 LSGIATFLMTIMIRIPAVKLEELSKTLDVFLVLPNHLGM-AVSGFYENYETRYCTSS 1257  
QY 612 E-----YLVKQGDLSFN-----GLWKNHVALACMIVFLTIAYLKLFL 651  
Db 1258 EVAHYCKYNIQIYQENFYANSAPGVGRFVSAAS-----GCAYLILULFL 1303

## RESULT 11

US-08-762-500-75  
; Sequence 75, Application US/08762500  
; Patent No. 6038606  
; GENERAL INFORMATION:  
; APPLICANT: Landes, Gregory M.  
; APPLICANT: Burn, Timothy C.  
; APPLICANT: Connors, Timothy D.  
; APPLICANT: Dackowski, William R.  
; APPLICANT: Van Raay, Terence J.  
; APPLICANT: Klinger, Katherine W.  
; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,  
; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME  
; NUMBER OF SEQUENCES: 83  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: GENZYME CORPORATION  
; STREET: One Mountain Road  
; CITY: Framingham  
; STATE: Massachusetts  
; COUNTRY: United States of America  
; ZIP: 01701  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/762,500  
; FILING DATE: 09-DEC-1996

CLASSIFICATION: 435  
PRIOR APPLICATION DATA: US 08/665,259  
FILING DATE: 17-JUN-1996  
PRIOR APPLICATION DATA: PCT/US96/10469  
FILING DATE: 17-JUN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Dugan, Deborah A.  
REGISTRATION NUMBER: 37,315  
REFERENCE/DOCKET NUMBER: IG5-9.3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (508) 872-8400  
TELEFAX: (508) 872-5415  
INFORMATION FOR SEQ ID NO: 75:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1704 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-762-500-75  
Query Match 7.1%; Score 238.5; DB 3; Length 1704;  
Best Local Similarity 20.3%; Pred. No. 8.2e-15;  
Matches 169; Conservative 116; Mismatches 241; Indels 305; Gaps 39;  
QY 66 NINGIMPGNALIGPTGGGKSSLLDLVLAARKDPSGLSGDVLING---APRPANFKNSG 122  
Db 553 NLN-LYEGQITVLLGHNGAGKTTLSMLTGLFPPT--SGRAYISGYEISQDMVQIRKSLG 609  
QY 123 YVQDDVVMGTLTVRENLFQSAALRLATTTNHEKNERINRVIOELGLDKVADSKVGTQF 182  
Db 610 LCPQHDILFNLTVAEHLFYAQLK---GLSRQKCPVEVKOMLHIGLEKWSNR-----661  
QY 183 IRGVSGERKRTSGMELITDPSILFDEPTTGLDSTANAVALLLKRMSKQGRITIFS 242  
Db 662 SRFLSGGRRKLSIGIALIAGSKVLILDEPTSGMDAISRRAIWDLQR--OKSDRTIVLTT 720  
QY 243 HQPRYSIFK---LPDSLTLASGRIMFHGPA---QEALGYFESAGYHC---EAYNPP 290  
Db 721 H-----FMDEADLLGDRJAJMAKGLQCGSSFLKQYK-----AGYHMTLVKEPHCNP 770  
QY 291 AD-----FFLD-----296  
Db 771 EDISQLVHHVHPNATLESSSAGAELSFILPRESTRPEGLFAKLEKKQKELGSIASFGASIT 830  
QY 297 -----IING 300  
Db 831 TMEEVFLRVKGLVDSSMDIQAIQLPALQYQHERRASDAVDSNLCGAMPDPSDGIGALIEE 890  
QY 301 DSTAVALN-----REEDFKATEIIEP-----SKQ- 324  
Db 891 ERTAVKLTGLALHCOQFWAMFLKKAAYSWEKRWAAQVLVPLTCVTLLALAINYSSEL 950  
QY 325 -DKPLI-----EKLAEE-----IYVNSSFFKETAELHQ-----351  
Db 951 FDPMLRLTLGEYGRVTVVPSVPGTSQQLQSEHLKDALQAEQEPREVLGDLEEFLLIF 1010  
QY 352 ---LSGGEKKKKTIVFKEISYTTSPCHQLRWKSKRSFKNLLGNPQ-----ASIAQI 399  
Db 1011 RASVEGGGFNERCLV-----AASF-----RDVGERTVNNALFNNQAYHSPATALAVVDNL 1060  
QY 400 IVTVVLGLVIGAYFGLKNDSTGIQ-----NRAGVLFFLTNNQCFSS-----SVSAV 445  
Db 1061 LFKLLCGPHASIVVSNFPQPSRSALQAQKQFNEGRKGFIALNLLFAMAFLASTFSILAV 1120  
QY 446 ELFVVEKKKLFTH-EYISGYRVSSYFGLKLLSDLEMTMLPSIIFTCIVVFMGLKPKAD 504  
Db 1121 SERAVQAK---HVQFVSGVH-VASFWSALLNDLISF-LIPSLLL--LVVF-----X 1165  
QY 505 AFFVMMFTLMVAVSASSMALATAAGOSVVSATLMTICFVFMIFSGLLVNLTIASW 564  
Db 1166 AFDVRAFT--RDGHMADTLLLLLYGWAIPLMYLNN-----FFFLGATATVTRLTIFI 1218



;; PRIOR APPLICATION NUMBER: US 60/055,779  
;; NUMBER OF SEQ ID NOS: 5674  
;; SEQ ID NO 3641  
;; LENGTH: 273  
;; TYPE: PR1  
;; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-3641

Query Match 6.8%; Score 227.5; DB 4; Length 273;  
Best Local Similarity 29.0%; Pred. No. 5.2e-15;  
Matches 76; Conservative 45; Mismatches 78; Indels 63; Gaps 13;  
QY 61 KEILSININGMKPG-LNAILGPTGGKSSLLDLVLAARKDPSGL-----SGDVINGAPRPA 115  
DB 31 KHVLEINIKINGEFLAIVPGNAGKSTLLKVL-----GLLPIQKGEIIVDGKP--- 81  
QY 116 NFKCNS-----GTVQ-----DDVVGTLTVRENJ--QPSAALRLATMTWH 155  
DB 82 -FKGNKSSLSKISVSQKASAFNAGFPASVKEVLSGLTKTKLQRF-----NK 129  
QY 156 EKNERINRVQELGLDKVADSKVGTQFIRGVSGGERKRTSIGMELITDPSILFLDEPTTG 215  
DB 130 NDQKVTKVLERNINSHLINKNAE-----LSGGQQRVLIAEALISDSVLVLEPTNG 184  
QY 216 LDSTANAVALLKRMKSGQRTIIFSIHQPRYSIFKLFDSLTLA--SGRLMFGPAQ-- 271  
DB 185 IDAKHVSFYDTLDKLLKQGITIILVTH-----DIGVVADTATEVACLKHLHFGHSTAP 240  
QY 272 EALGYEPESAGYHCEAYNNPADF 293  
DB 241 KSLDEVEIS-----KIYGHPIQF 258

## RESULT 15

US-08-232-537-2  
; Sequence 2, Application US/08232537  
; Patent No. 5516655  
; GENERAL INFORMATION:  
; APPLICANT: Peery, Robert B.  
; APPLICANT: Skatrud, Paul L.  
; TITLE OF INVENTION: MULTIPLE DRUG RESISTANCE GENE OF  
; TITLE OF INVENTION: AUREOBASIDIUM PULLULANS  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Eli Lilly and Company  
; STREET: Lilly Corporate Center  
; CITY: Indianapolis  
; STATE: Indiana  
; COUNTRY: USA  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,537  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plant, Thomas G.  
; REGISTRATION NUMBER: 35,784  
; REFERENCE/DOCKET NUMBER: X9212  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 317-276-2459  
; TELEFAX: 317-276-1917  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1302 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein

## US-08-232-537-2

Query Match 6.8%; Score 226.5; DB 1; Length 1302;  
Best Local Similarity 20.4%; Pred. No. 9.2e-14;  
Matches 139; Conservative 102; Mismatches 231; Indels 209; Gaps 27;  
QY 4 SNVEVFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICVRVKLSGFLPCRKPVEKEI 63  
DB 383 SEINVYSEAGQEA-----ESDMK-----ADIVFENVTF-----VYPAR--TSARA 421  
QY 64 LSNINGMKPG-LNAILGPTGGKSSLLDLVLAARKDPSGLSGDVINGAPRPAFKCNS- 121  
DB 422 LEEMSLIKAGOMNAIVGTSGCGKSTLVSLRLYDIS--SGQLTI--GSHDIKDFNVRSL 478  
QY 122 --GVYQDDVVMGTLTVRENLOPSAALRLATMTWHEKNERINRVQELGLDKV----- 173  
DB 479 RXYTALVDQSVLPSGVLNLSYGLG---EHSLSDDVLERCTEAKAANLQFVDFLPQ 535  
QY 174 -ADSKVGTQFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVALLKRM 232  
DB 536 GIHTRIGNGGYTSLSGGQNRICLARALVKKPALLLDDEPTAALDANSEGLINDAVKSA 595  
QY 233 KQRTIIFSIHQPRYSIFKLFDSLTLASGLRMFGHGAQAL---GYFESAGYHCEAYN 288  
DB 596 ATGTTVVNVAH--RLSTVSDSPNIVLMGAKVIEQGNHDELMQLEGAYFNL--LQAQLN 651  
QY 289 NPADFLDIINGDSTAVA---LNREDFKA---TEIIEP--SKQDKPLIEKLABIYVNSS 340  
DB 652 DADESSAENVSAATTSQVTPQKASKSDSAASSDTETVPPQAKKEDKA----- 699  
QY 341 FYKETKAELHOLSGEKKKTIYFKEISYTTSPCHQLRWVSKGSFKNLGNPQASIAQII 400  
DB 700 -----KKAGFWKLLRLCURLAKSDSPIALG-----LAASI 730  
QY 401 VTVVLGLVIG-AIYFGLKNDSTGIQNEAGVLFPLTTNQCFSSVSASVELFVVEKKLFIHEY 459  
DB 731 VSG--GIILGEAIVFG-----NLISVNLDESFD----- 757  
QY 460 ISGYRVSSYFLGKLLSLLPMTLPSIITCIYVFMGLKPKKADAFVMMFLMVA-- 517  
DB 758 -----FRSRADLPSLFFILALIALF 778  
QY 518 -YSASSMALAIAGQSVSVVATL-LMTICFVFMIFSG-----LLVNLTTIASWLSLQY 570  
DB 779 SYAGNGCCFGIVSSHFAKIQHISLASILRQDMQWFSQSQSVPSLSSSDAGQLCLSG 838  
QY 571 PSIPRYGFTALQHNFEFLQNPCLNATGNPCNYATCTGEEYLVKQIDLSFPLKXNH 630  
DB 839 VAI-----GTIFTVCVSTIG-----GIILAHVVAWKIA 866  
QY 631 VALACMIVIFLTIATVLLKLLFL 651  
DB 867 VLLAAVPMITAGYVRLVL 887

Search completed: March 23, 2004, 18:32:19  
Job time : 33 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2004, 18:23:18 ; Search time 60 Seconds  
(without alignments)  
3084.477 Million cell updates/sec

Title: US-09-961-086-1  
Perfect score: 3352  
Sequence: 1 MSSNVVEFIPVSGQTNGF.....MIVPLTIAYLKLLFLKYS 655

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

- 1: Geneseq1980s:\*
- 2: Geneseq1990s:\*
- 3: Geneseq2000s:\*
- 4: Geneseq2001s:\*
- 5: Geneseq2002s:\*
- 6: Geneseq2003as:\*
- 7: Geneseq2003bs:\*
- 8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3352	100.0	655	5	Aau80029 Human ABC
2	3352	100.0	663	2	Aay15221 Breast Ca
3	3346	99.8	655	4	Aab60104 Human tra
4	3346	99.8	655	5	Aao14781 Human BCR
5	3346	99.8	655	5	Aau80028 Human ABC
6	3346	99.8	655	6	Abr58077 Human ABC
7	3346	99.8	655	6	Ada10917 Human cdv
8	3346	99.8	655	7	Adc54182 Human bre
9	3339	99.6	665	5	Aao14783 Human BCR
10	3338	99.6	655	5	Abb07273 Human BCR
11	3331	99.4	655	3	Aay93365 ATP-bind
12	3331	99.4	655	4	Aau04348 Human BCR
13	3331	99.4	655	5	Abp52127 Homo sapi
14	3331	99.4	655	5	Abb07270 Human BCR
15	3331	99.4	655	6	Abu63376 Human mlt
16	3331	99.4	655	5	Aao14782 Human BCR
17	3053.5	91.1	604	2	Aaw73627 Human sec
18	3053.5	91.1	604	5	Abp61858 Human pol
19	2757	82.2	657	5	Abb07272 Murine BC
20	2325	69.4	456	4	Aab93564 Human pro
21	812	24.2	687	4	Abbs9384 Drosophil
22	808	24.1	687	3	Aay78981 Silkworm
23	793.5	23.4	625	3	Aag18080 Arabidops
24	783.5	23.4	632	3	Ang18079 Arabidops
25	783.5	23.4	648	3	Aag18078 Arabidops

26	746	22.3	649	3	AAG43100 Arabidops
27	746	22.3	653	3	AAG43099 Arabidops
28	727	21.7	832	4	ABBS9544 Arabidops
29	720	21.5	616	3	AAG50555 Drosophil
30	714.5	21.3	666	5	ABBS7112 Mouse isc
31	711	21.2	646	5	Aae28968 Human ABC
32	710.5	21.2	674	5	Abp52126 Homo sapi
33	710	21.2	638	5	Abb98349 Human ABC
34	710	21.2	646	5	Aae28964 Human ABC
35	710	21.2	646	5	Abb98348 Human ABC
36	708	21.1	627	6	Abbo07272 Human p53
37	708	21.1	646	5	Abb99694 Amino aci
38	707	21.1	646	5	Aao14186 Human tra
39	706.5	21.1	662	6	Abbo07271 Human p53
40	702.5	21.0	663	6	Abb82647 Human Dev
41	700	20.9	646	6	Abb99696 Amino aci
42	694.5	20.7	652	5	Aau96985 Mouse ABC
43	693.5	20.7	652	5	Aae13289 Mouse sit
44	693.5	20.7	652	5	Aae13289 Mouse sit
45	693.5	20.7	652	6	Aae31702 Mouse ABC

ALIGNMENTS

RESULT 1  
AAU80029  
ID AAU80029 standard; protein; 655 AA.  
AC AAU80029;  
XX  
DT 15-JUL-2002 (first entry)  
XX  
DE Human ABCG2 mutant 482T.  
XX  
KW Human; ABCG2; transporter protein; anticancer drug tolerance;  
KW indocarbazole; mutant; muten.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 482 /note= "Wild type Arg substituted by Thr"  
FT  
XX  
FN WO200228894-A1.  
XX  
PD 11-APR-2002.  
XX  
PF 18-SEP-2001; 2001WO-JP008112.  
XX  
PR 03-OCT-2000; 2000JP-00303441.  
XX  
PA (BANY ) BANYU PHARM CO LTD.  
XX  
PI Komatani H, Hara Y, Kotani H, Nakagawa R;  
XX  
DR WPI: 2002-352228/38.  
DR N-PSDB; ABK49911.  
XX  
PT ABCG2 gene encoding transporter protein capable of selectively  
PT transporting indocarbazole compounds, useful in screening inhibitors and  
PT anticancer agents for administration in chemotherapy.  
XX  
PS Disclosure; Page 87-90; 98pp; Japanese.  
XX  
CC The invention relates to an ABCG2 gene encoding a transporter protein  
CC capable of imparting tolerance to an anticancer agent in mammals  
CC comprising a fully defined sequence as given in the specification or an  
CC amino acid sequence based on the sequence but with some amino acids  
CC substituted, deleted or added. The gene and encoded protein are useful in  
CC screening inhibitors and anticancer agents for administration in  
CC chemotherapy with enhancement in sensitivity of cancer cell tolerance.

CC The gene relating to drug tolerance can be modified e.g. with the  
 CC transporter inhibitors, screened compounds, antibodies and antisense  
 CC nucleosides. The transporter is capable of selectively transporting  
 CC indocarbazole compounds extracellularly. The present sequence represents  
 CC the amino acid sequence of human ABCG2 mutant 482T  
 XX  
 SQ Sequence 655 AA;

Query Match 100.0%; Score 3352; DB 5; Length 655;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSSNNVEVPIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPYE 60  
 Db 1 MSSNNVEVPIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPYE 60

QY 61 KEILSNINGIMKPGNALIGPTGGKSSLLDVLARKDPSGLSDGLVINGAPRANFKCN 120  
 Db 61 KEILSNINGIMKPGNALIGPTGGKSSLLDVLARKDPSGLSDGLVINGAPRANFKCN 120

QY 121 SGYVQDDVVMGTLTVRENLFQSAALRLATTMTNHEKNERINRVIOELGDKVADSKVGT 180  
 Db 121 SGYVQDDVVMGTLTVRENLFQSAALRLATTMTNHEKNERINRVIOELGDKVADSKVGT 180

QY 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVILLKMSKQGRITIF 240  
 Db 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVILLKMSKQGRITIF 240

QY 241 SIHQPRYSIFKLFDSLTLIASGRMLFHGPAQEAALGYFESAGYHCEAYNPNADFLDIING 300  
 Db 241 SIHQPRYSIFKLFDSLTLIASGRMLFHGPAQEAALGYFESAGYHCEAYNPNADFLDIING 300

QY 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAETVYNSFVKETKAEHLQSGEKKKK 360  
 Db 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAETVYNSFVKETKAEHLQSGEKKKK 360

QY 361 ITVFKIEISYTSFCHQLRWKSKSPKLLGNPQASIAQIIVTVVLGLVIGAIYFGLKNDK 420  
 Db 361 ITVFKIEISYTSFCHQLRWKSKSPKLLGNPQASIAQIIVTVVLGLVIGAIYFGLKNDK 420

QY 421 TGIONRAGVLPFLTTNQCFSSVSAVELFVVEKLFIEHYISGYRVSSYFGLKLLDLP 480  
 Db 421 TGIONRAGVLPFLTTNQCFSSVSAVELFVVEKLFIEHYISGYRVSSYFGLKLLDLP 480

QY 481 MTMLPSIIPTCIVYFMLGLKPADAFFVMMFTLMVAYSASSMALAIAGQSVSVATLL 540  
 Db 481 MTMLPSIIPTCIVYFMLGLKPADAFFVMMFTLMVAYSASSMALAIAGQSVSVATLL 540

QY 541 MTICFVFMIFGLLVNLTIIASWLSWLOYSIPRYGFTALQHNFLQNGFCPGINATGN 600  
 Db 541 MTICFVFMIFGLLVNLTIIASWLSWLOYSIPRYGFTALQHNFLQNGFCPGINATGN 600

QY 601 NPCNATCTGEEYLVKQGDIDSPWGLKKNHVALACMIVIFLTIAVKLLFLKKYS 655  
 Db 601 NPCNATCTGEEYLVKQGDIDSPWGLKKNHVALACMIVIFLTIAVKLLFLKKYS 655

RESULT 2  
 AAY15221  
 ID AAY15221 standard; protein; 663 AA.  
 XX  
 AC AAY15221;  
 XX  
 XX 09-NOV-1999 (first entry)  
 DT Breast Cancer Resistance Protein (BCRP).  
 DE  
 XX breast cancer; drug resistance; ATP-binding cassette; ABC;  
 KW xenobiotic transporter; chemotherapy; mitoxantrone; doxorubicin;  
 KW breast cancer resistance protein; BCRP.  
 XX  
 OS Homo sapiens.  
 XX

FH Key Location/Qualifiers  
 FT Domain 87..95  
 FT /note= "Walker A motif"  
 FT Domain 221..236  
 FT /note= "Phosphotantetheine site"  
 FT Modified-site 345..347  
 FT /note= "Glycosylation site on N"  
 FT Region 405..422  
 FT /label= TM1  
 FT Modified-site 425..427  
 FT /note= "Transmembrane region"  
 FT Region 546..563  
 FT /label= TM2  
 FT Modified-site 564..566  
 FT /note= "Glycosylation site on N"  
 FT Modified-site 604..606  
 FT /note= "Glycosylation site on N"  
 FT Region 638..655  
 FT /label= TM3  
 XX WO9940110-A1.  
 PN 12-AUG-1999.  
 XX 05-FEB-1999; 99WO-US002577.  
 XX 05-FEB-1998; 98US-0073763P.  
 XX (UYMA-) UNIV MARYLAND BALTIMORE.  
 XX Ross DD, Doyle LA, Abruzzo L;  
 XX WPI; 1999-494273/41.  
 XX N-PSDB; AA206360.  
 XX New breast cancer resistance protein useful for production of antibodies  
 XX to inhibit resistance activity for enhancing chemotherapy treatment.  
 XX Claim 4; Fig 2a; 80pp; English.  
 XX The Breast Cancer Resistance Protein (BCRP) is an ATP-binding cassette  
 XX (ABC) transporter protein. It has a molecular mass of approximately 72.3  
 XX kilodaltons (kD) exclusive of any glycosylation. Expression of BCRP in  
 XX drug sensitive human cancer cells confers resistance to mitoxantrone, in  
 XX doxorubicin, and daunorubicin, and reduces daunorubicin accumulation in  
 XX the cloned transfected cells. The protein is useful for producing  
 XX antibodies and antisense probes, which can be used to inhibit the  
 XX activity of BCRP, therefore enhancing a cancer patient's chemotherapy  
 XX treatment. The antibodies and probes overcome the problems of breast  
 XX cancer resistance proteins to make chemotherapy treatment more effective  
 XX  
 SQ Sequence 663 AA;

Query Match 100.0%; Score 3352; DB 2; Length 663;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 655; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSSNNVEVPIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPYE 60  
 Db 9 MSSNNVEVPIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPYE 68

QY 61 KEILSNINGIMKPGNALIGPTGGKSSLLDVLARKDPSGLSDGLVINGAPRANFKCN 120  
 Db 69 KEILSNINGIMKPGNALIGPTGGKSSLLDVLARKDPSGLSDGLVINGAPRANFKCN 128

QY 121 SGYVQDDVVMGTLTVRENLFQSAALRLATTMTNHEKNERINRVIOELGDKVADSKVGT 180  
 Db 129 SGYVQDDVVMGTLTVRENLFQSAALRLATTMTNHEKNERINRVIOELGDKVADSKVGT 188

QY 181 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVILLKMSKQGRITIF 240  
 Db 189 QFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVILLKMSKQGRITIF 248

CC the diagnosis and treatment of transport, metabolic, neurological,  
 CC reproductive, cardiovascular and immune disorders, and cell proliferative  
 CC disorders such as cancer  
 XX  
 SQ Sequence 655 AA;

Query Match 99.8%; Score 3346; DB 4; Length 655;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSSNEVEVFIPVSGNTGPPATASNDLKAFTEGAVLSPHNICYRVKLSGFFPCPKPVE 60  
 DB 1 MSSNEVEVFIPVSGNTGPPATASNDLKAFTEGAVLSPHNICYRVKLSGFFPCPKPVE 60

QY 61 KEILSNINGIMKGLNALIGPTGGKSSLLDLAAKDPGLSGDVLINGAPPANFKCN 120  
 DB 61 KEILSNINGIMKGLNALIGPTGGKSSLLDLAAKDPGLSGDVLINGAPPANFKCN 120

QY 121 SGYVQDDVVMGTLTVRENIQFSAALRLATMTNHEKNERINRVIQELGLDKVADSKVGT 180  
 DB 121 SGYVQDDVVMGTLTVRENIQFSAALRLATMTNHEKNERINRVIQELGLDKVADSKVGT 180

QY 181 QFIRGVSGERKRTSGMELITDPSILFLDEPTTGLDSDSTANAVLLLLKEMSKOGRITIF 240  
 DB 181 QFIRGVSGERKRTSGMELITDPSILFLDEPTTGLDSDSTANAVLLLLKEMSKOGRITIF 240

QY 241 SIHQPRYSIFKLFDSLTLLASGRIMFHPGPAQEAALGYFESAGYHCEAYNNPADFFLDIING 300  
 DB 241 SIHQPRYSIFKLFDSLTLLASGRIMFHPGPAQEAALGYFESAGYHCEAYNNPADFFLDIING 300

QY 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEIYVNSSFYKETKAEHLQSGGEKKK 360  
 DB 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEIYVNSSFYKETKAEHLQSGGEKKK 360

QY 361 ITVFEKISYTTSFCHQLRWVSKRSFKNLGNPQASIAQIITVTVGLVIGALYFGLKXDS 420  
 DB 361 ITVFEKISYTTSFCHQLRWVSKRSFKNLGNPQASIAQIITVTVGLVIGALYFGLKXDS 420

QY 421 TGIONRAGVLPFLTNQCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFLGKLSDLLP 480  
 DB 421 TGIONRAGVLPFLTNQCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFLGKLSDLLP 480

QY 481 MTMLPSIIFTCIVVFMGLKPKADAFVMMFTLMVAYSASSMALAIAAGQSVSVATLL 540  
 DB 481 MTMLPSIIFTCIVVFMGLKPKADAFVMMFTLMVAYSASSMALAIAAGQSVSVATLL 540

QY 541 MTICFVFMIFPSGLLVNLTITIASWLSWLOYSIPRYGFTALQHNFLGQNFPCGLNATGN 600  
 DB 541 MTICFVFMIFPSGLLVNLTITIASWLSWLOYSIPRYGFTALQHNFLGQNFPCGLNATGN 600

QY 601 NPCNYATCTGEEYLVKQIDLSPWGLMKNHVALACMIVIFLTITAYLKLFLKKYS 655  
 DB 601 NPCNYATCTGEEYLVKQIDLSPWGLMKNHVALACMIVIFLTITAYLKLFLKKYS 655

RESULT 4  
 AAO14781  
 ID AAO14781 standard; protein; 655 AA.  
 XX  
 AC AAO14781;  
 XX  
 DT 28-JUN-2002 (first entry)  
 XX  
 DE Human BCRP protein.  
 XX  
 KW Human; BCRP protein; membrane penetrating region; cancer.  
 XX  
 OS Homo sapiens.  
 XX  
 PN JP2002065277-A.  
 XX  
 PD 05-MAR-2002.  
 XX

241 SIHQPRYSIFKLFDSLTLLASGRIMFHPGPAQEAALGYFESAGYHCEAYNNPADFFLDIING 300  
 DB 249 SIHQPRYSIFKLFDSLTLLASGRIMFHPGPAQEAALGYFESAGYHCEAYNNPADFFLDIING 308

QY 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEIYVNSSFYKETKAEHLQSGGEKKK 360  
 DB 309 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEIYVNSSFYKETKAEHLQSGGEKKK 368

QY 361 ITVFEKISYTTSFCHQLRWVSKRSFKNLGNPQASIAQIITVTVGLVIGALYFGLKXDS 420  
 DB 369 ITVFEKISYTTSFCHQLRWVSKRSFKNLGNPQASIAQIITVTVGLVIGALYFGLKXDS 428

QY 421 TGIONRAGVLPFLTNQCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFLGKLSDLLP 480  
 DB 429 TGIONRAGVLPFLTNQCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFLGKLSDLLP 488

QY 481 MTMLPSIIFTCIVVFMGLKPKADAFVMMFTLMVAYSASSMALAIAAGQSVSVATLL 540  
 DB 489 MTMLPSIIFTCIVVFMGLKPKADAFVMMFTLMVAYSASSMALAIAAGQSVSVATLL 548

QY 541 MTICFVFMIFPSGLLVNLTITIASWLSWLOYSIPRYGFTALQHNFLGQNFPCGLNATGN 600  
 DB 549 MTICFVFMIFPSGLLVNLTITIASWLSWLOYSIPRYGFTALQHNFLGQNFPCGLNATGN 608

QY 601 NPCNYATCTGEEYLVKQIDLSPWGLMKNHVALACMIVIFLTITAYLKLFLKKYS 655  
 DB 609 NPCNYATCTGEEYLVKQIDLSPWGLMKNHVALACMIVIFLTITAYLKLFLKKYS 663

RESULT 3  
 AAB60104  
 ID AAB60104 standard; protein; 655 AA.  
 XX  
 AC AAB60104;  
 XX  
 DT 28-MAR-2001 (first entry)  
 XX  
 DE Human transport protein TPPT-24.  
 XX  
 KW Human; transport protein; TPPT; transport disorder; metabolic disorder;  
 KW neurological disorder; cardiovascular disorder; reproductive disorder;  
 KW immune disorder; cancer.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200078953-A2.  
 XX  
 PD 28-DEC-2000.  
 XX  
 PF 16-JUN-2000; 2000WO-US016668.  
 XX  
 PR 17-JUN-1999; 99US-0139923P.  
 PR 10-AUG-1999; 99US-0148177P.  
 PR 18-AUG-1999; 99US-0149357P.  
 PR 28-OCT-1999; 99US-0162287P.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Lal P, Yang J, Yue H, Hillman JL, Tang YT, Bandman O, Burford N;  
 PI Baughn WR, Azimzai Y, Lu DAM, Au-Young J, Patterson C;  
 XX  
 DR WPI; 2001-041424/05.  
 DR N-PSDB; AAF27724.  
 XX  
 XX Isolated polypeptide with a human transport protein sequence is useful  
 PT for the diagnosis, prevention and treatment of disorders associated with  
 PT the immune, reproductive and cardiovascular systems.  
 XX  
 PS Claim 2; Page 126-127; 165pp; English.  
 XX  
 CC The present invention provides the protein and coding sequences for 43  
 CC novel human transport proteins (designated TPPTs). These can be used in



PF 31-AUG-2000; 2000JP-00263742.  
XX PR 31-AUG-2000; 2000JP-00263742.  
XX PA (GANK-) ZH GAN KENKYUKAI.  
XX WPI; 2002-324198/36.  
XX N-PSDB; AAL42412.  
XX PT Mutant BCRP protein useful for treatment of cancer.  
XX PS Claim 13; Page 7-8; 15pp; Japanese.  
XX CC The invention comprises a mutant human BCRP protein, having a deletion,  
CC replacement or addition of at least one amino acid in the fifth membrane  
CC penetrating region of the wild-type BCRP protein. The mutant BCRP protein  
CC can be used for the treatment of cancer. The present amino acid sequence  
CC represents a human BCRP protein  
XX SQ Sequence 655 AA;  
  
Query Match 99.8%; Score 3346; DB 5; Length 655;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 1 MSSNVEVFPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICVRVKLSGFLPCRKPE 60  
Db 1 MSSNVEVFPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICVRVKLSGFLPCRKPE 60  
  
Qy 61 KEILSNINGIMKPGNALGPTGGKSLDLVLAARKDPSGLSDVLINGAPRANFKCN 120  
Db 61 KEILSNINGIMKPGNALGPTGGKSLDLVLAARKDPSGLSDVLINGAPRANFKCN 120  
  
Qy 121 SGYVQDDVVMGTLTVRENLFSAALRLATTMTNHEKNERINRVIOELGLDKVADSKVGT 180  
Db 121 SGYVQDDVVMGTLTVRENLFSAALRLATTMTNHEKNERINRVIOELGLDKVADSKVGT 180  
  
Qy 181 QFIRVSGGERKRTSIGMELITDPSILFDEPTTGLDSSANAVLLLLKRMKSGQRTIIF 240  
Db 181 QFIRVSGGERKRTSIGMELITDPSILFDEPTTGLDSSANAVLLLLKRMKSGQRTIIF 240  
  
Qy 241 SIHQPRYSIFKLFDSLTLLASGRLMFHGPAQEALGFESAGYHCEAYNNPADPFLDIING 300  
Db 241 SIHQPRYSIFKLFDSLTLLASGRLMFHGPAQEALGFESAGYHCEAYNNPADPFLDIING 300  
  
Qy 301 DSTAVALNREEDPKATEIIEPSKQDKPLIEKLAIIYVNSFYKETAELHQLSGEKKKK 360  
Db 301 DSTAVALNREEDPKATEIIEPSKQDKPLIEKLAIIYVNSFYKETAELHQLSGEKKKK 360  
  
Qy 361 ITVFKIEISYTSFCHOLRWVSKSPKLLGNFQASIAQIIVTVVLGLVIGAIYFGLKND 420  
Db 361 ITVFKIEISYTSFCHOLRWVSKSPKLLGNFQASIAQIIVTVVLGLVIGAIYFGLKND 420  
  
Qy 421 TGIQNRAGVLPFLTNQCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFLGKLLSDLLP 480  
Db 421 TGIQNRAGVLPFLTNQCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFLGKLLSDLLP 480  
  
Qy 481 NPMPLSIIFTCTIVYFMGLKPKADAFVMMFTLMWVAYSASSMALAIAGOSVSVATLL 540  
Db 481 NPMPLSIIFTCTIVYFMGLKPKADAFVMMFTLMWVAYSASSMALAIAGOSVSVATLL 540  
  
Qy 541 MTICFVPMWIFSGLLVNLTTIASWLSWLOYSIPRYGFTALOHNEFLGQNCPCGLNATGN 600  
Db 541 MTICFVPMWIFSGLLVNLTTIASWLSWLOYSIPRYGFTALOHNEFLGQNCPCGLNATGN 600  
  
Qy 601 NPCNATCTGEBYLVKQGLDPSWGLKXNHVALACWIVFIATYKLLFLKKYS 655  
Db 601 NPCNATCTGEBYLVKQGLDPSWGLKXNHVALACWIVFIATYKLLFLKKYS 655

RESULT 5  
AAU80028  
ID AAU80028 standard; protein; 655 AA.

XX AC AAU80028;  
XX DT 15-JUL-2002 (first entry)  
XX DE Human ABCG2.  
XX KW Human; ABCG2; transporter protein; anticancer drug tolerance;  
XX KW indocarbazole.  
XX OS Homo sapiens.  
XX PN WO200228894-A1.  
XX PD 11-APR-2002.  
XX PP 18-SEP-2001; 2001WO-JP008112.  
XX PR 03-OCT-2000; 2000JP-00303441.  
XX PA (BANY ) BANYU PHARM CO LTD.  
XX PI Komatani H, Hara Y, Kotani H, Nakagawa R;  
XX WPI; 2002-352228/38.  
XX N-PSDB; ABK49901.  
XX ABCG2 gene encoding transporter protein capable of selectively  
XX transporting indocarbazole compounds, useful in screening inhibitors and  
XX anticancer agents for administration in chemotherapy.  
XX Claim 1; Page 71-76; 98pp; Japanese.  
XX The invention relates to an ABCG2 gene encoding a transporter protein  
XX capable of imparting tolerance to an anticancer agent in mammals  
XX comprising a fully defined sequence as given in the specification or an  
XX amino acid sequence based on the sequence but with some amino acids  
XX substituted, deleted or added. The gene and encoded protein are useful in  
XX screening inhibitors and anticancer agents for administration in  
XX chemotherapy with enhancement in sensitivity of cancer cell tolerance.  
XX The gene relating to drug tolerance can be modified e.g. with the  
XX transporter inhibitors, screened compounds, antibodies and antisense  
XX nucleotides. The transporter is capable of selectively transporting  
XX indocarbazole compounds extracellularly. The present sequence represents  
XX the amino acid sequence of human ABCG2 protein  
XX SQ Sequence 655 AA;  
  
Query Match 99.8%; Score 3346; DB 5; Length 655;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 1 MSSNVEVFPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICVRVKLSGFLPCRKPE 60  
Db 1 MSSNVEVFPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICVRVKLSGFLPCRKPE 60  
  
Qy 61 KEILSNINGIMKPGNALGPTGGKSLDLVLAARKDPSGLSDVLINGAPRANFKCN 120  
Db 61 KEILSNINGIMKPGNALGPTGGKSLDLVLAARKDPSGLSDVLINGAPRANFKCN 120  
  
Qy 121 SGYVQDDVVMGTLTVRENLFSAALRLATTMTNHEKNERINRVIOELGLDKVADSKVGT 180  
Db 121 SGYVQDDVVMGTLTVRENLFSAALRLATTMTNHEKNERINRVIOELGLDKVADSKVGT 180  
  
Qy 181 QFIRVSGGERKRTSIGMELITDPSILFDEPTTGLDSSANAVLLLLKRMKSGQRTIIF 240  
Db 181 QFIRVSGGERKRTSIGMELITDPSILFDEPTTGLDSSANAVLLLLKRMKSGQRTIIF 240  
  
Qy 241 SIHQPRYSIFKLFDSLTLLASGRLMFHGPAQEALGFESAGYHCEAYNNPADPFLDIING 300  
Db 241 SIHQPRYSIFKLFDSLTLLASGRLMFHGPAQEALGFESAGYHCEAYNNPADPFLDIING 300  
  
Qy 301 DSTAVALNREEDPKATEIIEPSKQDKPLIEKLAIIYVNSFYKETAELHQLSGEKKKK 360

Db 301 DSTAVLNREEDFKATEIIEPSKODKPLIEKLAIEYVNSFFYKETAELHQLSGGKKK 360  
Qy 361 ITVFKISIVTTSFCHOLRWVSKSPKMLGNPOASIAQIIIVTWGLVIGALYFGLKND 420  
Db 361 ITVFKISIVTTSFCHOLRWVSKSPKMLGNPOASIAQIIIVTWGLVIGALYFGLKND 420  
Qy 421 TGONRAGLVFLITTNOCFSSVAVELFVVEKKLFTHIEYISGYRVSSYFLGKLSDLLP 480  
Db 421 TGONRAGLVFLITTNOCFSSVAVELFVVEKKLFTHIEYISGYRVSSYFLGKLSDLLP 480  
Qy 481 MTMLPSIIFTCTIVYFVLMGLKPKADAFVVMFTLMWVAYSASSMALAIAAGQSVSVATLL 540  
Db 481 MTMLPSIIFTCTIVYFVLMGLKPKADAFVVMFTLMWVAYSASSMALAIAAGQSVSVATLL 540  
Qy 541 MTICFVFMIFSGLLVNLTTIASWLSWLOYSIPRYGFTALOHNEFLGONFCPLNATGN 600  
Db 541 MTICFVFMIFSGLLVNLTTIASWLSWLOYSIPRYGFTALOHNEFLGONFCPLNATGN 600  
Qy 601 NPCNYATCTGEEYLVKQIDLSFWGLWKNHVALACMIVIFLTIAYLKLLFLKKYS 655  
Db 601 NPCNYATCTGEEYLVKQIDLSFWGLWKNHVALACMIVIFLTIAYLKLLFLKKYS 655

RESULT 6  
ABR58077  
ID ABR58077 standard; protein; 655 AA.  
XX  
AC ABR58077;  
XX  
DT 15-OCT-2003 (first entry)  
XX  
DE Human ABCG2 protein.  
XX  
KW ABCG2; antidiabetic; cell therapy; diabetes mellitus;  
KW pancreatic stem cell; islets of langerhans; insulin.  
XX  
OS Homo sapiens.  
XX  
PN WO2003026584-A2.  
XX  
PD 03-APR-2003.  
XX  
XX 26-SEP-2002; 2002WO-US030700.  
XX  
XX 26-SEP-2001; 2001US-00963875.  
XX  
XX 11-APR-2002; 2002US-00120687.  
XX  
XX 02-MAY-2002; 2002US-00136891.  
XX  
XX (GEO ) GEN HOSPITAL CORP.  
XX  
XX Habener JF, Zulewski H, Thomas MK, Abraham EJ, Vallejo M;  
XX Leach CA, Nolan AL, Lechner A;  
XX  
XX WPI; 2003-354625/33.  
XX  
XX N-PSDB; ACC80605.  
XX

PT Treating a patient with diabetes mellitus by isolating a nestin- or ABCG2  
PT -positive pancreatic stem cell from a pancreatic islet of a donor and  
PT transferring the stem cell into the patient.  
XX  
XX Disclosure; Fig 18B; 107pp; English.  
XX  
XX The invention relates to a method of treating a patient with diabetes  
XX mellitus by isolating a nestin- or ABCG2-positive pancreatic stem cell  
XX from a pancreatic islet of a donor, and transferring the stem cell into  
XX the patient whereby the stem cell differentiates into an insulin-  
XX producing cell. Alternatively, the nestin- or ABCG2-positive stem is  
XX induced into a pancreatic progenitor cell prior to isolation and  
XX transfer. This sequence corresponds to the human ABCG2 protein and the  
XX encoding gene is detected in the method of the invention. The method is  
XX useful for preparing a pharmaceutical composition for treating diabetes  
XX mellitus. The stem cells can be further characterised for correct gene

CC expression using the primers and probes ACC80607-ACC80671  
XX Sequence 655 AA;  
Qy Query Match 99.8%; Score 3346; DB 6; Length 655;  
XX Best Local Similarity 99.8%; Pred. No. 0;  
Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 MSSNVVFIPVSGQNTNGFPATASNDLKAFTGAVLSFHNICVRVKLSGFLPCRKPE 60  
Db 1 MSSNVVFIPVSGQNTNGFPATASNDLKAFTGAVLSFHNICVRVKLSGFLPCRKPE 60  
Qy 61 KEILSNINGIMKPGNAILGPTGGKSKSLDLVAARKDPSGLSGDVLINGAPRANPKCN 120  
Db 61 KEILSNINGIMKPGNAILGPTGGKSKSLDLVAARKDPSGLSGDVLINGAPRANPKCN 120  
Qy 121 SGYVQDDVVGTLTVRENLOFSAALRLATMTNHEKNERINRVIOELGLDKVADSKVGT 180  
Db 121 SGYVQDDVVGTLTVRENLOFSAALRLATMTNHEKNERINRVIOELGLDKVADSKVGT 180  
Qy 181 QFIRGSGGERKRTSIGMELITDPSILFDBPTTGLDSSSTANAVLLLLKMSKQRTIIF 240  
Db 181 QFIRGSGGERKRTSIGMELITDPSILFDBPTTGLDSSSTANAVLLLLKMSKQRTIIF 240  
Qy 241 SIHQPRYSIFKLFDSLTLASGRMLMFGPAQALGYFESAGYHCEAYNNPADFFLDIING 300  
Db 241 SIHQPRYSIFKLFDSLTLASGRMLMFGPAQALGYFESAGYHCEAYNNPADFFLDIING 300  
Qy 301 DSTAVLNREEDFKATEIIEPSKODKPLIEKLAIEYVNSFFYKETAELHQLSGGKKK 360  
Db 301 DSTAVLNREEDFKATEIIEPSKODKPLIEKLAIEYVNSFFYKETAELHQLSGGKKK 360  
Qy 361 ITVFKISIVTTSFCHOLRWVSKSPKMLGNPOASIAQIIIVTWGLVIGALYFGLKND 420  
Db 361 ITVFKISIVTTSFCHOLRWVSKSPKMLGNPOASIAQIIIVTWGLVIGALYFGLKND 420  
Qy 421 TGONRAGLVFLITTNOCFSSVAVELFVVEKKLFTHIEYISGYRVSSYFLGKLSDLLP 480  
Db 421 TGONRAGLVFLITTNOCFSSVAVELFVVEKKLFTHIEYISGYRVSSYFLGKLSDLLP 480  
Qy 481 MTMLPSIIFTCTIVYFVLMGLKPKADAFVVMFTLMWVAYSASSMALAIAAGQSVSVATLL 540  
Db 481 MTMLPSIIFTCTIVYFVLMGLKPKADAFVVMFTLMWVAYSASSMALAIAAGQSVSVATLL 540  
Qy 541 MTICFVFMIFSGLLVNLTTIASWLSWLOYSIPRYGFTALOHNEFLGONFCPLNATGN 600  
Db 541 MTICFVFMIFSGLLVNLTTIASWLSWLOYSIPRYGFTALOHNEFLGONFCPLNATGN 600  
Qy 601 NPCNYATCTGEEYLVKQIDLSFWGLWKNHVALACMIVIFLTIAYLKLLFLKKYS 655  
Db 601 NPCNYATCTGEEYLVKQIDLSFWGLWKNHVALACMIVIFLTIAYLKLLFLKKYS 655

RESULT 7  
ADA10917  
ID ADA10917 standard; protein; 655 AA.  
XX  
AC ADA10917;  
XX  
DT 06-NOV-2003 (first entry)  
XX  
DE Human cDNA differentially expressed in colon cancer #23 product.  
XX  
XX differential expression; colon cancer; cancer; human.  
XX  
XX Homo sapiens.  
XX  
XX US2002160392-A1.  
XX  
XX 31-OCT-2002.  
XX  
XX 11-OCT-2001; 2001US-00981353.  
XX

```

PR 11-OCT-2000; 2000US-0239841P.
XX (LASE/) LASEK A W.
PA (JONE/) JONES D A.
XX Lasek AW, Jones DA;
FI
DR WPI; 2003-265756/26.
DR N-PSDB; ADA10916.
XX
PT New combination comprising cDNAs that are differentially expressed in
PT colon disorder, useful for diagnosing, treating, staging or monitoring
PT treatment for colon cancer.
XX
PS Example 14; SEQ ID NO 35; 231pp; English.
XX
CC The invention relates to a combination comprising cDNAs that are
CC differentially expressed in colon disorder. The methods and compositions
CC of the present invention are useful for diagnosing, treating, staging or
CC monitoring treatment for colon cancer. They are also useful in high
CC throughput methods for using cDNAs to detect differential expression of
CC nucleic acids in a sample, screening molecules or compounds to identify a
CC ligand which specifically binds a cDNA and using a protein to screen
CC molecules or compounds to identify at least one ligand which specifically
CC binds the protein. The present sequence represents the amino acid
CC sequence of a human cDNA differentially expressed in colon cancer
CC protein.
XX
SQ Sequence 655 AA;
Query Match 99.8%; Score 3346; DB 6; Length 655;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MSSNVVEVFPVSOQNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPYE 60
Db 1 MSSNVVEVFPVSOQNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPYE 60
QY 61 KEILSNINGIMKPGNALIGPTGGKSSLLDVLARKDPSGLSGDVLINGAPRANFKCN 120
Db 61 KEILSNINGIMKPGNALIGPTGGKSSLLDVLARKDPSGLSGDVLINGAPRANFKCN 120
QY 121 SGYVQDDVVMGTLVRENLOPSAALRLATTWNHEKNERINRVIOELGLDKVADSKVGT 180
Db 121 SGYVQDDVVMGTLVRENLOPSAALRLATTWNHEKNERINRVIOELGLDKVADSKVGT 180
QY 181 QIRGVSGGERKRTSIGMELITDPSILFDLDEPTTGLDSTANAVLLLLKRMKSGKRTIIF 240
Db 181 QIRGVSGGERKRTSIGMELITDPSILFDLDEPTTGLDSTANAVLLLLKRMKSGKRTIIF 240
QY 241 SIHQPRYSIFKLFDSLTLASGRMLPHGPAQALGYFESAGYHCEAYNNPADFFLDIING 300
Db 241 SIHQPRYSIFKLFDSLTLASGRMLPHGPAQALGYFESAGYHCEAYNNPADFFLDIING 300
QY 301 DSTAVALNEEDFKATEIIEPSKQDKPLEKLAIEYVNSSFYKETAELHOLSGEKKKK 360
Db 301 DSTAVALNEEDFKATEIIEPSKQDKPLEKLAIEYVNSSFYKETAELHOLSGEKKKK 360
QY 361 ITVFKESYTSFCHQLRWVSKSRFNKLLGNPQASIAQIIVTVLGLVIGAIYFGLKND 420
Db 361 ITVFKESYTSFCHQLRWVSKSRFNKLLGNPQASIAQIIVTVLGLVIGAIYFGLKND 420
QY 421 TGIONPAGVLFLTNNOCFSSYSAVELFVVEKKLFTHIYISGYEVSSYFGLKLLDLP 480
Db 421 TGIONPAGVLFLTNNOCFSSYSAVELFVVEKKLFTHIYISGYEVSSYFGLKLLDLP 480
QY 481 NTMLPSIIFTCIVYFMGLKPKADAFFVMMFTLMWVAYSASSMALAIAAGQSVSVATLL 540
Db 481 NTMLPSIIFTCIVYFMGLKPKADAFFVMMFTLMWVAYSASSMALAIAAGQSVSVATLL 540
QY 541 MTICFVMMIFSGLLVNLITIASLSWLOQYFIPRYGFTALOHNEFLGONPCPLNATGN 600
Db 541 MTICFVMMIFSGLLVNLITIASLSWLOQYFIPRYGFTALOHNEFLGONPCPLNATGN 600
QY 601 NPCNYATCTGBEYLVKQIGIDLSPGMLWKNHVALACMIVIFLTIAYLKLLFLKKYS 655
Db 601 NPCNYATCTGBEYLVKQIGIDLSPGMLWKNHVALACMIVIFLTIAYLKLLFLKKYS 655
RESULT 8
ADC54182
ID ADC54182 standard; protein; 655 AA.
XX
AC ADC54182;
XX
DT 18-DEC-2003 (first entry)
XX
DE Human breast cancer resistance protein (BCRP) amino acid sequence.
XX
KW cancer cell; anti-cancer agent; steroid hormone; oestrogenic effect;
KW BCRP; breast cancer resistance protein; cytostatic; camptothecins;
KW mitoxantrone; 7-hydroxy staurosporine; adriamycin; cancer chemotherapy;
KW human.
XX
OS Homo sapiens.
XX
PN JP2003063989-A.
XX
PD 05-MAR-2003.
XX
PF 23-AUG-2001; 2001JP-00252953.
XX
PR 23-AUG-2001; 2001JP-00252953.
XX
PA (GANK-) ZH GAN KENYUKAI.
XX
DR WPI; 2003-735321/70.
XX
N-PSDB; ADC54181.
PT Agent that overcomes resistance of cancer cell against anti-cancer agent,
PT comprises a steroid hormone, or a compound which exhibits antagonistic
PT activity against the hormone, with the cancer cell expressing BCRP gene.
XX
PS Example 1; SEQ ID NO 4; 15pp; Japanese.
XX
CC This invention relates to a novel agent which overcomes resistance of a
CC cancer cell against an anti-cancer agent (AA), comprising as an active
CC ingredient a steroid hormone, a compound having oestrogenic effect, or a
CC compound which exhibits antagonistic activity against the hormone, where
CC the cancer cell expresses the BCRP (breast cancer resistance protein)
CC gene. The agent of the invention may have cytostatic activity. The
CC invention is useful for overcoming resistance of a cancer against an anti
CC -cancer agent such as camptothecin, mitoxantrone, 7-hydroxy
CC staurosporine and adriamycin. The therapeutic effective anti-cancer agent
CC is recovered, due to the use of the agent of the invention. Also the
CC dosages of anti-cancer agent can be maintained easily, and adverse
CC effects of cancer chemotherapy can be suppressed. The present sequence is
CC that of the human BCRP protein which was used to develop the novel agent
CC of the invention.
XX
SQ Sequence 655 AA;
Query Match 99.8%; Score 3346; DB 7; Length 655;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MSSNVVEVFPVSOQNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPYE 60
Db 1 MSSNVVEVFPVSOQNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPYE 60
QY 61 KEILSNINGIMKPGNALIGPTGGKSSLLDVLARKDPSGLSGDVLINGAPRANFKCN 120
Db 61 KEILSNINGIMKPGNALIGPTGGKSSLLDVLARKDPSGLSGDVLINGAPRANFKCN 120
QY 121 SGYVQDDVVMGTLVRENLOPSAALRLATTWNHEKNERINRVIOELGLDKVADSKVGT 180
Db 121 SGYVQDDVVMGTLVRENLOPSAALRLATTWNHEKNERINRVIOELGLDKVADSKVGT 180

```

Db 121 SGVVQDDVVMGTLTVRENLQFSAALRLATTWTHKEKNERINRVIOELGLDKVADSKVGT 180  
 Qy 181 QFIRGVSGGERKRTSIGMELITDPSILFDLDEPTTGLDSDSTANAVLLLLKRMKSKQGRITIF 240  
 Db 181 QFIRGVSGGERKRTSIGMELITDPSILFDLDEPTTGLDSDSTANAVLLLLKRMKSKQGRITIF 240  
 Qy 241 SIHQPRYSIFKLFDSLTLLASGRLMFHGPAQALGYFESAGVHCEAYNNPADFFLDIING 300  
 Db 241 SIHQPRYSIFKLFDSLTLLASGRLMFHGPAQALGYFESAGVHCEAYNNPADFFLDIING 300  
 Qy 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAETIYVNSSFYKETAELHQLSGGKCKK 360  
 Db 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAETIYVNSSFYKETAELHQLSGGKCKK 360  
 Qy 361 ITVFKETSYTTSFCHOLRWKSKSFKNLLGNPOASIAQIIVTVLGLVIGAIYFGLKNDK 420  
 Db 361 ITVFKETSYTTSFCHOLRWKSKSFKNLLGNPOASIAQIIVTVLGLVIGAIYFGLKNDK 420  
 Qy 421 TGIQNRAGVLFLLTNQCFSSVSAVELFVVEKKLFIHEIYISGYRVSSYFGLKLLSDLLP 480  
 Db 421 TGIQNRAGVLFLLTNQCFSSVSAVELFVVEKKLFIHEIYISGYRVSSYFGLKLLSDLLP 480  
 Qy 481 MTWLPSSIIFTCIYVFMGLGPKADAFVMMFTLMVAYSSASSMALAIAGQSVSVATLL 540  
 Db 481 MRMLPSIIFTCIYVFMGLGPKADAFVMMFTLMVAYSSASSMALAIAGQSVSVATLL 540  
 Qy 541 MTICFVFMIFSGLLVNLTTIASWLSWLOQFSIPRYGFTALQHNFLGQNFPCPLNATGN 600  
 Db 541 MTICFVFMIFSGLLVNLTTIASWLSWLOQFSIPRYGFTALQHNFLGQNFPCPLNATGN 600  
 Qy 601 NPCNYATCTGEEYLVKQGDIDSPWGLWKNHVALACMIVIFLTAYLKLFLKKYS 655  
 Db 601 NPCNYATCTGEEYLVKQGDIDSPWGLWKNHVALACMIVIFLTAYLKLFLKKYS 655

## RESULT 9

AAO14783  
 ID AAO14783 standard; protein; 665 AA.

XX AAO14783;

XX 28-JUN-2002 (first entry)

XX Human BCRP-related protein 2.

XX Human; BCRP-related protein; membrane penetrating region; cancer.

XX Homo sapiens.

XX JP2002065277-A.

XX 05-MAR-2002.

XX 31-AUG-2000; 2000JP-00263742.

XX 31-AUG-2000; 2000JP-00263742.

XX (GANK-) ZH GAN KENYUKAI.

XX WPI; 2002-324198/36.

XX N-PSDB; AAL42414.

XX Mutant BCRP protein useful for treatment of cancer.

XX Example 3; Page 12-13; 15pp; Japanese.

XX The invention comprises a mutant human BCRP protein, having a deletion, replacement or addition of at least one amino acid in the fifth membrane penetrating region of the wild-type BCRP protein. The mutant BCRP protein can be used for the treatment of cancer. The present amino acid sequence represents a human BCRP-related protein

XX Sequence 665 AA;

PR 31-MAY-2000; 2000US-00584586.

Query Match 99.6%; Score 3339; DB 5; Length 665;  
 Best Local Similarity 99.7%; Pred. No. 0;  
 Matches 653; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MSSSNEVEFIPVSGQNTNGFPATASNDLKAFTEGAVLSFPHNICYRVKLSGFLPCRKPVE 60  
 Db 11 MSSSNEVEFIPVSGQNTNGFPATASNDLKAFTEGAVLSFPHNICYRVKLSGFLPCRKPVE 70  
 Qy 61 KEILSNINGIMKPLGNAIIGPTGGKSSLLDLAARKDPSGLSGDVLINGARPPANFKCN 120  
 Db 71 KEILSNINGIMKPLGNAIIGPTGGKSSLLDLAARKDPSGLSGDVLINGARPPANFKCN 130  
 Qy 121 SGVVQDDVVMGTLTVRENLQFSAALRLATTWTHKEKNERINRVIOELGLDKVADSKVGT 180  
 Db 131 SGVVQDDVVMGTLTVRENLQFSAALRLATTWTHKEKNERINRVIOELGLDKVADSKVGT 190  
 Qy 181 QFIRGVSGGERKRTSIGMELITDPSILFDLDEPTTGLDSDSTANAVLLLLKRMKSKQGRITIF 240  
 Db 191 QFIRGVSGGERKRTSIGMELITDPSILFDLDEPTTGLDSDSTANAVLLLLKRMKSKQGRITIF 250  
 Qy 241 SIHQPRYSIFKLFDSLTLLASGRLMFHGPAQALGYFESAGVHCEAYNNPADFFLDIING 300  
 Db 251 SIHQPRYSIFKLFDSLTLLASGRLMFHGPAQALGYFESAGVHCEAYNNPADFFLDIING 310  
 Qy 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAETIYVNSSFYKETAELHQLSGGKCKK 360  
 Db 311 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAETIYVNSSFYKETAELHQLSGGKCKK 370  
 Qy 361 ITVFKETSYTTSFCHOLRWKSKSFKNLLGNPOASIAQIIVTVLGLVIGAIYFGLKNDK 420  
 Db 371 ITVFKETSYTTSFCHOLRWKSKSFKNLLGNPOASIAQIIVTVLGLVIGAIYFGLKNDK 430  
 Qy 421 TGIQNRAGVLFLLTNQCFSSVSAVELFVVEKKLFIHEIYISGYRVSSYFGLKLLSDLLP 480  
 Db 431 TGIQNRAGVLFLLTNQCFSSVSAVELFVVEKKLFIHEIYISGYRVSSYFGLKLLSDLLP 490  
 Qy 481 MTWLPSSIIFTCIYVFMGLGPKADAFVMMFTLMVAYSSASSMALAIAGQSVSVATLL 540  
 Db 491 MRMLPSIIFTCIYVFMGLGPKADAFVMMFTLMVAYSSASSMALAIAGQSVSVATLL 550  
 Qy 541 MTICFVFMIFSGLLVNLTTIASWLSWLOQFSIPRYGFTALQHNFLGQNFPCPLNATGN 600  
 Db 551 MTICFVFMIFSGLLVNLTTIASWLSWLOQFSIPRYGFTALQHNFLGQNFPCPLNATGN 610  
 Qy 601 NPCNYATCTGEEYLVKQGDIDSPWGLWKNHVALACMIVIFLTAYLKLFLKKYS 655  
 Db 611 NPCNYATCTGEEYLVKQGDIDSPWGLWKNHVALACMIVIFLTAYLKLFLKKYS 665

## RESULT 10

ABB07273

ID ABB07273 standard; protein; 655 AA.

XX ABB07273;

XX 26-MAR-2002 (first entry)

XX Human BCRP (huBCRP) sequence.

XX Stem cell; ATP transport protein; ATP-binding cassette; antiparkinsonian; hepatotropic; neurodegenerative; cytostatic; antianemic; muscular; BCRP; KW cardiant; gene therapy.

XX Homo sapiens.

XX WO200192877-A2.

XX 06-DEC-2001.

XX 30-MAY-2001; 2001WO-US017459.

XX 31-MAY-2000; 2000US-00584586.

29-MAY-2001; 2001US-00866866.  
 (STUD-) ST JUDE CHILDREN'S RES HOSPITAL.  
 Sorrentino B, Schuetz J;  
 WPI; 2002-114368/15.  
 N-PSDB; ABA94383.  
 Identifying a stem cell, for treating e.g., muscular dystrophy,  
 myocardial infarction, Parkinson's disease, or neurodegenerative  
 disorders, comprises detecting the expression of an ATP transport protein  
 (BCRP) by a cell.  
 Example 1; Page 86-87; 87pp; English.  
 The invention provides a method of identifying and/or isolating a stem  
 cell that involves detecting the expression of an ATP transport protein  
 containing a conserved ATP-binding cassette (ABCP) by a cell in a sample  
 comprising stem cells. The isolated stem cells may be used in the  
 treatment of diseases such as muscular dystrophy, degenerative liver  
 disorder, myocardial infarction, Parkinson's disease, degenerative  
 disorders of the brain, and for tissue regeneration or replacement.  
 Haematopoietic cells can be used in bone marrow transplants (e.g., for  
 treatment of leukemia) and for ex vivo gene therapy for treating blood  
 diseases such as sickle cell anemia and thalassemia. The stem cells can  
 also be used as cell targets in gene therapy protocols. The present  
 sequence represents the human BCRP  
 Sequence 655 AA;  
 Query Match . 99.6%; Score 3338; DB 5; Length 655;  
 Best Local Similarity 99.7%; Pred. No. 0;  
 Matches 653; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MSSNVEVFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICVRVKLSGFLPCRPVE 60  
 Db 1 MSSNVEVFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICVRVKLSGFLPCRPVE 60  
 QY 61 KEILSNINGIMKPGNALIGPTGGKSSLLDLVLAARKDPSGLSDVLINGAPRANFKCN 120  
 Db 61 KEILSNINGIMKPGNALIGPTGGKSSLLDLVLAARKDPSGLSDVLINGAPRANFKCN 120  
 QY 121 SGYVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIOELGLDKVADSKVGT 180  
 Db 121 SGYVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIOELGLDKVADSKVGT 180  
 QY 181 QFIRGVSGERKRTSIGMELITDPSILFLDEPTTGLDSDSTANAVILLKRMKSGRTIIF 240  
 Db 181 QFIRGVSGERKRTSIGMELITDPSILFLDEPTTGLDSDSTANAVILLKRMKSGRTIIF 240  
 QY 241 SIHQPRYSIFKFLDSLTLLASGRLMFHGPAQALGYFESAGVHCEAYNNPADFFDIING 300  
 Db 241 SIHQPRYSIFKFLDSLTLLASGRLMFHGPAQALGYFESAGVHCEAYNNPADFFDIING 300  
 QY 301 DSTAVALNDEEFKATEIIEPSKQDKPLEKLAETVWNSFVKETKAEHLHOLSGEKKKX 360  
 Db 301 DSTAVALNDEEFKATEIIEPSKQDKPLEKLAETVWNSFVKETKAEHLHOLSGEKKKX 360  
 QY 361 ITVFEKISYTTSFCHQLRWVSKRSFKNLGNPQASIAQIIVTVLGLVIGAYIFGLKNDS 420  
 Db 361 ITVFEKISYTTSFCHQLRWVSKRSFKNLGNPQASIAQIIVTVLGLVIGAYIFGLKNDS 420  
 QY 421 TGIONRAGVLFLTITNQCESSYSAVELFVVEKKLFIEHVISGYRVSSYFELGLSDLLP 480  
 Db 421 TGIONRAGVLFLTITNQCESSYSAVELFVVEKKLFIEHVISGYRVSSYFELGLSDLLP 480  
 QY 481 MTMLPSTIITFCIIVYFMGLKPKADAFVVMFTIMVAYSASSMALAIAGQSVSWATLL 540  
 Db 481 MTMLPSTIITFCIIVYFMGLKPKADAFVVMFTIMVAYSASSMALAIAGQSVSWATLL 540  
 QY 541 MTICFVFMIFPGLLVNLTITIASWLSWLOQYFSIPRYGFTALQHNELGQNFQGLNATGN 600

Db 541 MTICFVFMIFPGLLVNLTITIASWLSWLOQYFSIPRYGFTALQHNELGQNFQGLNATGN 600  
 QY 601 NPCNYATCTGEEYLVKQIDLSLSPWGLWKNHVALACMIVIFLTITAYLKLFLKKYS 655  
 Db 601 NPCNYATCTGEEYLVKQIDLSLSPWGLWKNHVALACMIVIFLTITAYLKLFLKKYS 655

## RESULT 11

AA95365  
 ID AA95365 standard; protein; 655 AA.

XX AC AA95365;

DT 25-SEP-2000 (first entry)

XX DE ATP-binding cassette protein MXR1.

XX KW ATP-binding cassette protein; ABC protein; MXR1; human;  
 cytotoxin resistance; colon carcinoma; therapy.

OS Homo sapiens.

XX PN WO200036101-A2.

XX PD 22-JUN-2000.

XX PF 24-NOV-1999; 99WO-US028107.

XX PR 30-NOV-1998; 98US-0110473P.

XX PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX PI Dean M, Allikmets R, Bates SE, Fojo AT;

XX DR WPI; 2000-442385/38.

XX DR N-PSDB; AAR27938.

XX PT Adenosine triphosphate (ATP)-binding cassette proteins and their nucleic  
 acids, useful for identifying agents that may be used to treat colonic  
 carcinomas resistant to cytotoxic therapeutic agents.

XX PS Claim 1(ii); Page 46; 49pp; English.

XX CC The present sequence is that of an ATP-binding cassette protein (ABC  
 protein) designated MXR1 that confers mitoxantrone resistance to SI-M1-80  
 human colon carcinoma cells. Nucleic acids encoding the ABC protein may  
 be used to produce the protein according to standard recombinant DNA  
 methodologies. The expressed proteins may be used to confer resistance to  
 cytotoxins such as mitoxantrone and/or daunomycin to the host cells in  
 which they are expressed. In disease conditions such as carcinoma of the  
 colon, breast and gastrointestinal tract, the ABC protein may be  
 expressed to provide resistance to cytotoxic therapeutic agents. The  
 nucleic acids and proteins may be used to produce host cell models of  
 resistant cells which can be used to screen for candidate agents that  
 inhibit the expression and/or activity of the ABC protein. These agents  
 may be used to down regulate ABC protein expression in carcinoma cells  
 and sensitize them to cytotoxic therapeutic agents

XX SQ Sequence 655 AA;

Query Match 99.4%; Score 3331; DB 3; Length 655;

Best Local Similarity 99.4%; Pred. No. 0;

Matches 651; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSSNVEVFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICVRVKLSGFLPCRPVE 60

Db 1 MSSNVEVFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICVRVKLSGFLPCRPVE 60

QY 61 KEILSNINGIMKPGNALIGPTGGKSSLLDLVLAARKDPSGLSDVLINGAPRANFKCN 120

Db 61 KEILSNINGIMKPGNALIGPTGGKSSLLDLVLAARKDPSGLSDVLINGAPRANFKCN 120

QY 121 SGYVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIOELGLDKVADSKVGT 180

```
Db 121 SGYVQDDVVGTLTVRENLQFSAALRLATTTNHEKNERINRVIEELGLDKVADSKVGT 180
Qy 181 QFIRGVSGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVALLKRMKSKQGRITIF 240
Db 181 QFIRGVSGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVALLKRMKSKQGRITIF 240
Qy 241 SIHQPRYSIFKLFDSLTLTLLASGRMLMFGPAQALGYFESAGYHCEAYNPNADFFLDIING 300
Db 241 SIHQPRYSIFKLFDSLTLTLLASGRMLMFGPAQALGYFESAGYHCEAYNPNADFFLDIING 300
Qy 301 DSTAVALNREEDFKATEIIEPSKODKPLIEKLAIEIYVNSFFYKETKAEHLHOLSGEKKKK 360
Db 301 DSTAVALNREEDFKATEIIEPSKODKPLIEKLAIEIYVNSFFYKETKAEHLHOLSGEKKKK 360
Qy 361 ITVFKEISYTTSFCHQLRWWSKRSFKNLLGNPQASIAQIIVTVVLGLVIGIYFGLKNDK 420
Db 361 ITVFKEISYTTSFCHQLRWWSKRSFKNLLGNPQASIAQIIVTVVLGLVIGIYFGLKNDK 420
Qy 421 TGQONRAGVLFLTTNQCFSSVSAVELFVVEKKLFHEIYISGYRVSSYFLGKLSDDLPL 480
Db 421 TGQONRAGVLFLTTNQCFSSVSAVELFVVEKKLFHEIYISGYRVSSYFLGKLSDDLPL 480
Qy 481 MTMLPSIIFTCIYVFMGLGPKADAPFVMFTLMVAYSASSMALAIAAGQSVSVATLL 540
Db 481 MRMLPSIIFTCIYVFMGLGPKADAPFVMFTLMVAYSASSMALAIAAGQSVSVATLL 540
Qy 541 MTICFVFMFTPSGLLVNLTIASWLMLOYESTPRYGTALQHNELGONFCPLGNATGN 600
Db 541 MTICFVFMFTPSGLLVNLTIASWLMLOYESTPRYGTALQHNELGONFCPLGNATGN 600
Qy 601 NPCNYATCTGEEYLVKQIDLSWGLWQHVALACMIVIFLTAYLKLFLKKYS 655
Db 601 NPCNYATCTGEEYLVKQIDLSWGLWQHVALACMIVIFLTAYLKLFLKKYS 655

RESULT 12
AAU04348
ID AAU04348 standard; protein; 655 AA.
AC AAU04348;
XX
DT 23-OCT-2001 (first entry)
XX
DE Human BCRP/MXR/ABCP protein.
XX
KW ATP-binding cassette transporter protein-inhibiting peptide; ABC;
KW transmembrane domain; cancer; tumour; HIV; AIDS; BCRP/MXR/ABCP;
KW acquired immunodeficiency syndrome; human immunodeficiency syndrome.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 398..420
FT FT /label= Transmembrane_domain_1
FT Domain 427..450
FT FT /label= Transmembrane_domain_2
FT Domain 478..501
FT FT /label= Transmembrane_domain_3
FT Domain 504..527
FT FT /label= Transmembrane_domain_4
FT Domain 539..558
FT FT /label= Transmembrane_domain_5
FT Domain 629..650
FT FT /label= Transmembrane_domain_6
XX
XX WO200136477-A2.
XX
XX 25-MAY-2001.
XX
XX 17-NOV-2000; 2000WO-US031817.
XX
XX 13-NOV-1999; 99US-0166382P.
```

```
PR 22-NOV-1999; 99US-0166767P.
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX Tarasova NI, Michejda CJ, Gottesman MM, Hrycyna C;
XX WPI; 2001-381224/40.
XX New ATP-binding cassette transporter-inhibiting peptides derived from
XX transmembrane domain of the transporter, useful for inhibiting biological
XX activity of the transporter and sensitizing cancer cells to
XX chemotherapeutics.
XX Disclosure; Fig 2; 89pp; English.
XX The sequence represents human BCRP/MXR/ABCP protein, an ATP-binding
XX cassette (ABC) transporter protein. Peptides derived from its
XX transmembrane domains are used as ATP-binding cassette transporter
XX protein-inhibiting peptides. The peptides have a first end and a second
XX end, where at the first end there is a group that is negatively charged
XX under physiological conditions and at the second end a group that is
XX neutrally charged under physiological conditions and is at least 70%
XX identical to a transmembrane domain of an ABC transporter. The peptides
XX are used for inhibiting the biological activity, in particular ion flux
XX or translocation, cytotoxin efflux or translocation, phosphorylation,
XX protein synthesis or degradation, cellular morphology, secretion,
XX production of particular components such as soluble inositol phosphates,
XX tumour growth, chemotaxis, mitogenic response, cell growth activation or
XX secretion of a target ATP-binding cassette (ABC) transporter. The
XX peptides are useful to sensitize cancer cells to standard
XX chemotherapeutics, which chemotherapeutics are then administered to kill
XX the cancer cells. Transmembrane analogues of ABC transporter protein are
XX administered to patients infected with HIV-1, who are taking a protease
XX inhibitor. The ability of HIV-1 infected cells to efflux the protease
XX inhibitor is impaired, resulting in higher concentration of the protease
XX inhibitor in the affected cell. This reduces the amount of protease
XX inhibitor administered to the patient to achieve a virus-inhibiting dose
XX in the infected cells
XX
XX SQ Sequence 655 AA;
XX
XX Query Match 99.4%; Score 3331; DB 4; Length 655;
XX Best Local Similarity 99.4%; Pred. No. 0;
XX Matches 651; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
Qy 1 MSSNVVEFIPVSGQNTGFPATASNDLKAPTEGAVLSFHNICYRVKLSGFLPCRPVE 60
Db 1 MSSNVVEFIPVSGQNTGFPATVNDLKAPTEGAVLSFHNICYRVKLSGFLPCRPVE 60
Qy 61 KEILSNINGIMKPGNALIGPTGGKSSLLDVLAAKDPGLSGDLVINGAPRANFKCN 120
Db 61 KEILSNINGIMKPGNALIGPTGGKSSLLDVLAAKDPGLSGDLVINGAPRANFKCN 120
Qy 121 SGYVQDDVVGTLTVRENLQFSAALRLATTTNHEKNERINRVIEELGLDKVADSKVGT 180
Db 121 SGYVQDDVVGTLTVRENLQFSAALRLATTTNHEKNERINRVIEELGLDKVADSKVGT 180
Qy 181 QFIRGVSGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVALLKRMKSKQGRITIF 240
Db 181 QFIRGVSGERKRTSIGMELITDPSILFLDEPTTGLDSTANAVALLKRMKSKQGRITIF 240
Qy 241 SIHQPRYSIFKLFDSLTLTLLASGRMLMFGPAQALGYFESAGYHCEAYNPNADFFLDIING 300
Db 241 SIHQPRYSIFKLFDSLTLTLLASGRMLMFGPAQALGYFESAGYHCEAYNPNADFFLDIING 300
Qy 301 DSTAVALNREEDFKATEIIEPSKODKPLIEKLAIEIYVNSFFYKETKAEHLHOLSGEKKKK 360
Db 301 DSTAVALNREEDFKATEIIEPSKODKPLIEKLAIEIYVNSFFYKETKAEHLHOLSGEKKKK 360
Qy 361 ITVFKEISYTTSFCHQLRWWSKRSFKNLLGNPQASIAQIIVTVVLGLVIGIYFGLKNDK 420
Db 361 ITVFKEISYTTSFCHQLRWWSKRSFKNLLGNPQASIAQIIVTVVLGLVIGIYFGLKNDK 420
```

421 TGIONRAGVLPFLTTNOCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFLGKLLSDLLP 480  
 Db 421 TGIONRAGVLPFLTTNOCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFLGKLLSDLLP 480  
 Qy 481 NTMLPSIIFTCTIVFMGLGKPKADAFVMMFTLMMVAYSASSMALAIAGQSVSVATLL 540  
 Db 481 NTMLPSIIFTCTIVFMGLGKPKADAFVMMFTLMMVAYSASSMALAIAGQSVSVATLL 540  
 Qy 541 MTICFVFMIFSGLLVNLTTIASLSWLQYFSIPRYGFTALQHNFFLGQFCPLNATGN 600  
 Db 541 MTICFVFMIFSGLLVNLTTIASLSWLQYFSIPRYGFTALQHNFFLGQFCPLNATGN 600  
 Qy 601 NPCNATCTGSEYLVKQIDISPMGLMKHVALACMIVIFLTAYLKLFLKKYS 655  
 Db 601 NPCNATCTGSEYLVKQIDISPMGLMKHVALACMIVIFLTAYLKLFLKKYS 655

## RESULT 13

ABP52127  
ID ABP52127 standard; protein; 655 AA.

XX AC ABP52127;

XX DT 10-OCT-2002 (first entry)

XX DE Homo sapiens ABC transporter ABCG2 protein SEQ ID NO:79.

XX KW ATP-binding cassette transporter; ABC transporter; modulation; D loop;  
 cancer; bacterial infection; fungal infection; protozoal infection;  
 antibacterial; fungicide; protozoaside.

XX OS Homo sapiens.

XX PN EP1217066-A1.

XX PD 26-JUN-2002.

XX PF 21-DEC-2000; 2000EP-00870316.

XX PR 21-DEC-2000; 2000EP-00870316.

XX PA (UYGE-) UNIV GENT.

XX DR WPI; 2002-550404/59.

XX PT Modulating activity of ATP-binding cassette (ABC) transporters by  
 influencing dimerization of nucleotide binding domains through use of D  
 loop sequence of an ABC transporter, or its antisense peptide or peptide  
 mimetic.

XX PS Disclosure; Fig 3; 290pp; English.

XX CC The present invention describes a method (M1) for modulating the activity  
 of ATP-binding cassette (ABC) transporters by influencing the  
 dimerization of the nucleotide binding domains comprising: (a) a  
 polypeptide (polyP) consisting of 5-50 amino acids comprising the D loop  
 sequence of an ABC transporter (ABP52049 to ABP52091); (b) a polyP  
 consisting of the D loop sequence of an ABC transporter; (c) a peptide  
 mimetic or antisense peptide of (a) or (b). ABC transporters have  
 antibacterial, fungicide and protozoaside activities. (M1) is useful for  
 selectively modulating the activity of ABC transporters belonging to the  
 group of multidrug transporter/P-glycoproteins. Bacterial, fungal or  
 protozoal ABC transporters are involved in the infection of a mammal or  
 in the induction of resistance to antibiotics or drugs in a mammal. (M1)  
 is useful for preventing, treating or alleviating diseases associated  
 with functionality of an ABC transporter. ABP52092 to ABP52140 represent  
 ABC transporter proteins given in the exemplification of the present  
 invention

XX SQ Sequence 655 AA;

Query Match 99.4%; Score 3331; DB 5; Length 655;  
 Best Local Similarity 99.4%; Pred. No. 0;

Qy	Matches	651;	Conservative	1;	Mismatches	3;	Indels	0;	Gaps	0;
1	MSSSNVEVFI	PVSQNTNGFP	PATASNDL	KAFTEG	AVLSFHN	ICYRVK	LKSGFL	PCRPVE	60	
1	MSSSNVEVFI	PVSQNTNGFP	PATVSN	DLKAFTEG	AVLSFHN	ICYRVK	LKSGFL	PCRPVE	60	
61	KEILSNING	IMKPGN	AILGPT	GGKSS	LLDVL	AARD	PSGL	SGDVL	INGAPR	ANFKCN 120
61	KEILSNING	IMKPGN	AILGPT	GGKSS	LLDVL	AARD	PSGL	SGDVL	INGAPR	ANFKCN 120
121	SGYVVD	DDVM	GTLV	REN	LQF	SAAL	RLATT	TNTN	HEKNER	INRV
121	SGYVVD	DDVM	GTLV	REN	LQF	SAAL	RLATT	TNTN	HEKNER	INRV
181	QFIRV	SGGER	KRTS	IGMEL	ITD	PSIL	FDE	PTT	GLD	SSAN
181	QFIRV	SGGER	KRTS	IGMEL	ITD	PSIL	FDE	PTT	GLD	SSAN
241	SIHQ	PRYS	IF	KL	FDS	LT	LLAS	GR	LM	FH
241	SIHQ	PRYS	IF	KL	FDS	LT	LLAS	GR	LM	FH
301	DSTAV	ALN	REED	FKATE	II	EPS	KQD	KPL	IE	KLA
301	DSTAV	ALN	REED	FKATE	II	EPS	KQD	KPL	IE	KLA
361	ITV	KEI	S	Y	T	T	S	F	C	H
361	ITV	KEI	S	Y	T	T	S	F	C	H
421	TGION	RAG	V	L	F	L	T	T	N	Q
421	TGION	RAG	V	L	F	L	T	T	N	Q
481	MTML	PSI	I	F	T	C	I	V	F	M
481	MTML	PSI	I	F	T	C	I	V	F	M
541	MTIC	F	V	F	M	I	F	S	G	L
541	MTIC	F	V	F	M	I	F	S	G	L
601	NPCN	AT	C	T	G	S	E	Y	L	V
601	NPCN	AT	C	T	G	S	E	Y	L	V

RESULT 14  
 ABB07270  
 ID ABB07270 standard; protein; 655 AA.  
 XX AC ABB07270;  
 XX DT 26-MAR-2002 (first entry)  
 XX DE Human BCRP (hBCRP) amino acid sequence.  
 XX KW Stem cell; ATP transport protein; ATP-binding cassette; antiparkinsonian;  
 hepatotropic; neurodegenerative; cytostatic; antianemic; muscular; BCRP;  
 cardiant; gene therapy.  
 XX OS Homo sapiens.  
 XX PN W0200192877-A2.  
 XX PD 06-DEC-2001.  
 XX PF 30-MAY-2001; 2001WO-US017459.  
 XX PR 31-MAY-2000; 2000US-00584586.  
 XX PR 29-MAY-2001; 2001US-00868866.  
 XX PA (SUJD-) ST JUDE CHILDREN'S RES HOSPITAL.



XX Sorrentino B, Schuetz J;  
XX WPI; 2002-114368/15.  
DR N-PSDB; ABA94369.  
XX Identifying a stem cell, for treating e.g., muscular dystrophy,  
PT myocardial infarction, Parkinson's disease, or neurodegenerative  
PT disorders, comprises detecting the expression of an ATP transport protein  
PT (BCRP) by a cell.  
XX Example; Page 78-80; 87pp; English.  
XX The invention provides a method of identifying and/or isolating a stem  
CC cell that involves detecting the expression of an ATP transport protein  
CC containing a conserved ATP-binding cassette (BCRP) by a cell in a sample  
CC comprising stem cells. The isolated stem cells may be used in the  
CC treatment of diseases such as muscular dystrophy, degenerative liver  
CC disorder, myocardial infarction, Parkinson's disease, degenerative  
CC disorders of the brain, and for tissue regeneration or replacement.  
CC Haematopoietic cells can be used in bone marrow transplants (e.g., for  
CC treatment of leukemia) and for ex vivo gene therapy for treating blood  
CC diseases such as sickle cell anemia and thalassemia. The stem cells can  
CC also be used as cell targets in gene therapy protocols. The present  
CC sequence represents the amino acid sequence of human BCRP  
XX Sequence 655 AA;  
SQ  
Query Match 99.4%; Score 3331; DB 5; Length 655;  
Best Local Similarity 99.4%; Pred. No. 0;  
Matches 651; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 1 MSSSNVEVFIPVSGQNTGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPE 60  
Db 1 MSSSNVEVFIPVSGQNTGFPATVSNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPE 60  
QY- 61 KEILSNINGIMKGLNAILGPTGGKSSLLDLVAARKDPSGLDVLINGARPRANFCN 120  
Db 61 KEILSNINGIMKGLNAILGPTGGKSSLLDLVAARKDPSGLDVLINGARPRANFCN 120  
QY 121 SGYVQDDVVMGTLTVRENIQFSAALRLATTMTNHEKNERINRVIOELGLDKVADSKVT 180  
Db 121 SGYVQDDVVMGTLTVRENIQFSAALRLATTMTNHEKNERINRVIEELGLDKVADSKVT 180  
QY 181 QFIRGVSGGRKRTSIGMELITDPSILFDEPTGLDSTANAVLLLLKMSKQGTIIF 240  
Db 181 QFIRGVSGGRKRTSIGMELITDPSILFDEPTGLDSTANAVLLLLKMSKQGTIIF 240  
QY 241 SIHQPRYSIFKLPDLSLTLLASGRIMFHPGPAQEAALGYFESAGYHCEAYNNPADPFLDIING 300  
Db 241 SIHQPRYSIFKLPDLSLTLLASGRIMFHPGPAQEAALGYFESAGYHCEAYNNPADPFLDIING 300  
QY 301 DSTAVALNRBEDFKATEIIBPSKQDKPLIEKLAETVYNSFYKETKAEHLQSLGSGKKK 360  
Db 301 DSTAVALNRBEDFKATEIIBPSKQDKPLIEKLAETVYNSFYKETKAEHLQSLGSGKKK 360  
QY 361 ITVPKEISYTFSPCHLRWYSKSFKNLLGNPQASIAQIIVTVVLGIVGAIYFGLKND 420  
Db 361 ITVPKEISYTFSPCHLRWYSKSFKNLLGNPQASIAQIIVTVVLGIVGAIYFGLKND 420  
QY 421 TGIQNRAGVLFFLTNTQCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFLGKLLSDLLP 480  
Db 421 TGIQNRAGVLFFLTNTQCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFLGKLLSDLLP 480  
QY 481 MTMLPSIIFTCIVVFMGLKPKADAFVVMFTLMMVAYSASSMALAIAGQSVSVATLL 540  
Db 481 MTMLPSIIFTCIVVFMGLKPKADAFVVMFTLMMVAYSASSMALAIAGQSVSVATLL 540  
QY 541 MTICFVFMIFSGLLVNLTTIASLWLYPSIPRYGFTALQHNFEFGQFCPLNATGN 600  
Db 541 MTICFVFMIFSGLLVNLTTIASLWLYPSIPRYGFTALQHNFEFGQFCPLNATGN 600  
QY 601 NPNYATCTGEEYLVKQIDLSLPGWGLKKNHVALACMIVIFLITAYLKLFLKKYS 655

Db 601 NPNYATCTGEEYLVKQIDLSLPGWGLKKNHVALACMIVIFLITAYLKLFLKKYS 655  
RESULT 15  
ABU63376  
ID ABU63376 standard; protein; 655 AA.  
XX AC ABU63376;  
XX 23-SEP-2003 (first entry)  
XX Human mitoxantrone resistance (MXR) transporter.  
XX Human; cytostatic; acridine derivative; MXR; Pgp; cancer;  
KW mitoxantrone resistance transporter; p-glycoprotein; lung cancer;  
KW colon cancer; breast cancer; prostate cancer; ovarian cancer;  
KW acute lymphocytic leukaemia; non-Hodgkin's lymphoma;  
KW multidrug resistance; MDR.  
XX Homo sapiens.  
XX US6521635-B1.  
XX 18-FEB-2003.  
XX 22-JAN-2001; 2001US-00767594.  
XX 20-JAN-2000; 2000US-0177410P.  
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.  
XX Bates S, Robey R;  
XX WPI; 2003-553619/52.  
XX New use of acridine derivative for inhibiting mitoxantrone resistance  
PT transporter in treatment of e.g. cancer, acute lymphocytic leukemia and  
PT non-Hodgkin's lymphoma.  
XX Disclosure; Col 29-34; 27pp; English.  
XX The invention relates to the use of an acridine derivative, or its salt  
CC or solvate, for inhibiting a mitoxantrone resistance (MXR) transporter in  
CC a cell over-expressing MXR gene but not over-expressing a p-glycoprotein  
CC (Pgp) gene, by contacting with the cell. Also included are assaying the  
CC modulation of the functional effect of a test compound on the cell by the  
CC acridine derivative (comprising contacting the test compound with the  
CC cells in the presence and absence of the acridine derivative and  
CC measuring the ability of the acridine derivative to modulate the  
CC functional effect of the test compound) and the treatment of cancer that  
CC over-expresses the (MXR) gene but not over-expresses the (Pgp) gene  
CC (comprising co-administering a chemotherapeutic recognized by the MXR  
CC transporter and the acridine derivative. The method is used for  
CC inhibiting the MXR transporter, and for treating cancer (e.g. lung  
CC cancer, colon cancer, breast cancer, prostate cancer, acute lymphocytic  
CC leukaemia, non-Hodgkin's lymphoma and ovarian cancer). The acridine  
CC derivative is a potent multispecific antagonist capable of inhibiting or  
CC reversing both Pgp-mediated and MXR-mediate multidrug resistance (MDR)  
CC phenotype. The present sequence is the human MXR transporter sequence  
XX Sequence 655 AA;  
SQ  
Query Match 99.4%; Score 3331; DB 6; Length 655;  
Best Local Similarity 99.4%; Pred. No. 0;  
Matches 651; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 1 MSSSNVEVFIPVSGQNTGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPE 60  
Db 1 MSSSNVEVFIPVSGQNTGFPATVSNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPE 60  
QY 61 KEILSNINGIMKGLNAILGPTGGKSSLLDLVAARKDPSGLDVLINGARPRANFCN 120

Db 61 KEILSNINGIMKPGNAILGPTGGKSSLLDVLAAKDPGSLGSDVLINGAPRANFKCN 120  
Qy 121 SGYVQDDVNVGTLTVRENLOFSAALRLATTMTNHEKNERINRVIOELGIDKVDKSVGT 180  
Db 121 SGYVQDDVNVGTLTVRENLOFSAALRLATTMTNHEKNERINRVIEELGIDKVDKSVGT 180  
Qy 181 QPIRGVSGGERKRTSIGNELITDPSILFLDEPTTGLDSTANAVLLLLKRMKSQGRITIF 240  
Db 181 QPIRGVSGGERKRTSIGNELITDPSILFLDEPTTGLDSTANAVLLLLKRMKSQGRITIF 240  
Qy 241 SIHOPRYISIFKLPDSLTLLASGRMLFHGPAQALGYFESAGYHCEAYNNPADPFLDIING 300  
Db 241 SIHOPRYISIFKLPDSLTLLASGRMLFHGPAQALGYFESAGYHCEAYNNPADPFLDIING 300  
Qy 301 DSTAVALNREEDFKATEIIEPSKODKPLIEKLAELIYVNSSFYKETKAELHQLSGGEKKK 360  
Db 301 DSTAVALNREEDFKATEIIEPSKODKPLIEKLAELIYVNSSFYKETKAELHQLSGGEKKK 360  
Qy 361 ITVFKEISYTTSFCHQLRWVSKRSFKNLLGNPOASIAQIIVTVVLGLVIGAIYFGLKND 420  
Db 361 ITVFKEISYTTSFCHQLRWVSKRSFKNLLGNPOASIAQIIVTVVLGLVIGAIYFGLKND 420  
Qy 421 TGIQNRAGVLPFLTTNQCFSSVSAVELFWVEKLFIEHYISGYRVSSYFLGKLLSDLLP 480  
Db 421 TGIQNRAGVLPFLTTNQCFSSVSAVELFWVEKLFIEHYISGYRVSSYFLGKLLSDLLP 480  
Qy 481 MTMLPSIITPTCIVYFMLGLPKADAFFVMMFTLMMVAYSASSMALAIAAGOSVVSATLL 540  
Db 481 MTMLPSIITPTCIVYFMLGLPKADAFFVMMFTLMMVAYSASSMALAIAAGOSVVSATLL 540  
Qy 541 MTICFVFMWIFSGLLVNLTTIASWSLQYFSIPRYGFTALQHNEFLGONFCPGLNATGN 600  
Db 541 MTICFVFMWIFSGLLVNLTTIASWSLQYFSIPRYGFTALQHNEFLGONFCPGLNATGN 600  
Qy 601 NPCNYATCTGEEYLVKQIGIDLSFWGLWKNHVALACMIVIFTIAYLKLFLKKYS 655  
Db 601 NPCNYATCTGEEYLVKQIGIDLSFWGLWKNHVALACMIVIFTIAYLKLFLKKYS 655

Search completed: March 23, 2004, 18:29:34  
Job time : 64 secs